ADAPTIVE ESTIMATION OVER ANISOTROPIC FUNCTIONAL CLASSES VIA ORACLE APPROACH

Oleg Lepski

The Annals of Statistics
2015, Vol. 43, No. 3, 1178–1242
DOI: 10.1214/14-AOS1306

ADAPTIVE ESTIMATION OVER ANISOTROPIC FUNCTIONAL CLASSES VIA ORACLE APPROACH

BY OLEG LEPSKI

Aix-Marseille Université

We address the problem of adaptive minimax estimation in white Gaussian noise models under $\mathbb{L}_p$-loss, $1 \le p \le \infty$, on the anisotropic Nikol'skii classes. We present the estimation procedure based on a new data-driven selection scheme from the family of kernel estimators with varying bandwidths. For the proposed estimator we establish so-called $\mathbb{L}_p$-norm oracle inequality and use it for deriving minimax adaptive results. We prove the existence of rate-adaptive estimators and fully characterize behavior of the minimax risk for different relationships between regularity parameters and norm indexes in definitions of the functional class and of the risk. In particular some new asymptotics of the minimax risk are discovered, including necessary and sufficient conditions for the existence of a uniformly consistent estimator. We provide also a detailed overview of existing methods and results and formulate open problems in adaptive minimax estimation.

1. Introduction. Let $\mathbb{R}^d, d \ge 1$, be equipped with Borel σ-algebra $\mathfrak{B}(\mathbb{R}^d)$ and Lebesgue measure ν_d. Put $\widetilde{\mathfrak{B}}(\mathbb{R}^d) = \{B \in \mathfrak{B}(\mathbb{R}^d) : \nu_d(B) < \infty\}$, and let $(W(B), B \in \widetilde{\mathfrak{B}}(\mathbb{R}^d))$ be the white noise with intensity ν_d. Set also for any $\mathcal{A} \in \mathfrak{B}(\mathbb{R}^d)$ and any $1 \le p < \infty$,

$$\mathbb{L}_p(\mathcal{A}, \nu_d) = \left\{ g : \mathcal{A} \to \mathbb{R} : \|g\|_{p,\mathcal{A}}^p := \int_{\mathcal{A}} |g(t)|^p \nu_d(\mathrm{d}t) < \infty \right\};$$

$$\mathbb{L}_\infty(\mathcal{A}) = \left\{ g : \mathcal{A} \to \mathbb{R} : \|g\|_{\infty,\mathcal{A}} := \sup_{t \in \mathcal{A}} |g(t)| < \infty \right\}.$$

1.1. *Statistical model and $\mathbb{L}_p$-risk.* Consider the sequence of statistical experiments (called Gaussian white noise models) generated by the observation $X^\varepsilon = \{X_\varepsilon(g), g \in \mathbb{L}_2(\mathbb{R}^d, \nu_d)\}_\varepsilon$ where

$$X_\varepsilon(g) = \int f(t) g(t) \nu_d(\mathrm{d}t) + \varepsilon \int g(t) W(\mathrm{d}t). \tag{1.1}$$

Here $\varepsilon \in (0, 1)$ is understood as the noise level which is usually supposed sufficiently small.

Received April 2014; revised November 2014.
MSC2010 subject classifications. 62G05, 62G20.
Key words and phrases. White Gaussian noise model, oracle inequality, adaptive estimation, kernel estimators with varying bandwidths, $\mathbb{L}_p$-risk.

The goal is to recover unknown signal f from observation X^ε on a given cube $(-b, b)^d$, $b > 0$. The quality of an estimation procedure will be described by $\mathbb{L}_p$-risk, $1 \le p \le \infty$, defined in (1.2) below, and as an estimator we understand any X^ε-measurable Borel function belonging to $\mathbb{L}_p(\mathbb{R}^d, \nu_d)$. Without loss of generality and for ease of notation, we will assume that functions to be estimated vanish outside $(-b, b)^d$.

Thus, for any estimator $\tilde{f}_\varepsilon$ and any $f \in \mathbb{L}_p(\mathbb{R}^d, \nu_d) \cap \mathbb{L}_2(\mathbb{R}^d, \nu_d)$, we define its $\mathbb{L}_p$-risk as

$$\mathcal{R}_\varepsilon^{(p)}[\tilde{f}_\varepsilon; f] = \{\mathbb{E}_f^{(\varepsilon)}(\|\tilde{f}_\varepsilon - f\|_p^q)\}^{1/q}, \qquad q \ge 1. \tag{1.2}$$

Here and throughout the paper, $\|\cdot\|_p$, $1 \le p \le \infty$ stands for $\|\cdot\|_{p,(-b,b)^d}$, and $\mathbb{E}_f^{(\varepsilon)}$ denote the mathematical expectation with respect to the probability law of X^ε.

Let $\mathbb{F}$ be a given subset of $\mathbb{L}_p(\mathbb{R}^d, \nu_d) \cap \mathbb{L}_2(\mathbb{R}^d, \nu_d)$. For any estimator $\tilde{f}_\varepsilon$ define its *maximal risk* by $\mathcal{R}_\varepsilon^{(p)}[\tilde{f}_\varepsilon; \mathbb{F}] = \sup_{f \in \mathbb{F}} \mathcal{R}_\varepsilon^{(p)}[\tilde{f}_\varepsilon; f]$ and its *minimax risk* on $\mathbb{F}$ is given by

$$\phi_\varepsilon(\mathbb{F}) := \inf_{\tilde{f}_\varepsilon} \mathcal{R}_\varepsilon^{(p)}[\tilde{f}_\varepsilon; \mathbb{F}]. \tag{1.3}$$

Here infimum is taken over all possible estimators. The estimator $\hat{f}$ is called minimax on $\mathbb{F}$ if

$$\limsup_{\varepsilon \to 0} \phi_\varepsilon^{-1}(\mathbb{F}) \mathcal{R}_\varepsilon^{(p)}[\hat{f}_\varepsilon; \mathbb{F}] < \infty.$$

1.2. *Adaptive estimation.* Let $\{\mathbb{F}_\vartheta, \vartheta \in \Theta\}$ be the collection of subsets of $\mathbb{L}_p(\mathbb{R}^d, \nu_d) \cap \mathbb{L}_2(\mathbb{R}^d, \nu_d)$, where ϑ is a nuisance parameter which may have very complicated structure.

The problem of adaptive estimation can be formulated as follows: *is it possible to construct a single estimator $\hat{f}_\varepsilon$ which would be simultaneously minimax on each class $\mathbb{F}_\vartheta$, $\vartheta \in \Theta$, that is,*

$$\limsup_{\varepsilon \to 0} \phi_\varepsilon^{-1}(\mathbb{F}_\vartheta) \mathcal{R}_\varepsilon^{(p)}[\hat{f}_\varepsilon; \mathbb{F}_\vartheta] < \infty \qquad \forall \vartheta \in \Theta?$$

We refer to this question as *the problem of adaptive estimation over the scale of* $\{\mathbb{F}_\vartheta, \vartheta \in \Theta\}$. If such estimator exists, we will call it optimally or rate-adaptive.

In the present paper we will be interested in adaptive estimation over the scale

$$\mathbb{F}_\vartheta = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \qquad \vartheta = (\vec{\beta}, \vec{r}, \vec{L}),$$

where $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ is an anisotropic Nikol'skii class; see Section 3.1 for a formal definition. Here we only mention that for any $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$, the coordinate β_i of the vector $\vec{\beta} = (\beta_1, \dots, \beta_d) \in (0, \infty)^d$ represents the smoothness of f in the direction i, and the coordinate r_i of the vector $\vec{r} = (r_1, \dots, r_d) \in [1, \infty]^d$ represents the index of the norm in which β_i is measured. Moreover, $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ is the intersection of the balls in some semi-metric space, and the vector $\vec{L} \in (0, \infty)^d$ represents the radii of these balls.

The aforementioned dependence on the direction is usually referred to *anisotropy* of the underlying function and the corresponding functional class. The use of the integral norm in the definition of the smoothness is referred to *inhomogeneity* of the underlying function. The latter means that the function f can be sufficiently smooth on some part of the observation domain and rather irregular on the other part. Thus the adaptive estimation over the scale $\{\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), (\vec{\beta}, \vec{r}, \vec{L}) \in (0, \infty)^d \times [1, \infty]^d \times (0, \infty)^d\}$ can be viewed as the adaptation to anisotropy and inhomogeneity of the function to be estimated.

1.3. *Historical notes*. The history of the adaptive estimation over scales of sets of smooth functions dates back 30 years. During this time a variety of functional classes was introduced in nonparametric statistics, particularly those of Sobolev, Nikol'skii and Besov. The relations between different scales as well as between classes belonging to the same scale can be found, for instance, in Nikol'skiĭ (1977). It is worth mentioning that although the considered classes are different, the same estimation procedure may be (nonadaptively) minimax on them. In such situations we will say that the class $\mathbb{F}_1$ is *statistically equivalent* to the class $\mathbb{F}_2$ and write $\mathbb{F}_1 \bowtie \mathbb{F}_2$. Also for two sequences $a_\varepsilon \to 0$ and $b_\varepsilon \to 0$, we will write $a_\varepsilon \sim b_\varepsilon$ and $a_\varepsilon \gtrsim b_\varepsilon$ if $0 < \lim_{\varepsilon\to 0} a_\varepsilon b_\varepsilon^{-1} < \infty$ and $\lim_{\varepsilon\to 0} a_\varepsilon b_\varepsilon^{-1} \geq 1$, respectively.

Estimation of univariate functions. The first adaptive results were obtained in Efroĭmovich and Pinsker (1984). The authors studied the problem of adaptive estimation over the scale of periodic Sobolev classes (Sobolev ellipsoids), $W(\beta, L)$, in the univariate model (1.1) under $\mathbb{L}_2$-loss ($p = 2$). The exact asymptotics of minimax risk on $W(\beta, L)$ is given by $P(L)\varepsilon^{2\beta/(2\beta+1)}$, where $P(L)$ is the Pinsker constant. The authors proposed the estimation procedure based on blockwise Bayesian construction and showed that it is adaptive and *efficient* over the scale of considered classes. Noting that $W(\beta, L) \bowtie \mathbb{N}_{2,1}(\beta, L)$, one can assert that Efroimovich–Pinsker estimator is rate-adaptive on $\mathbb{N}_{2,1}(\beta, L)$ as well.

Influenced by this pioneering paper, a variety of adaptive methods under $\mathbb{L}_2$-loss were proposed in different statistical models such as density and spectral density estimation, nonparametric regression, deconvolution model, inverse problems and many others. Let us mention some of them.

- Extension of the Efroimovich–Pinsker method, Efroimovich (1986, 2008);
- Unbiased risk minimization, Golubev (1992), Golubev and Nussbaum (1992);
- Model selection, Barron, Birgé and Massart (1999), Birgé (2008), Birgé and Massart (2001);
- Aggregation of estimators, Nemirovski (2000), Juditsky and Nemirovski (2000), Wegkamp (2003), Tsybakov (2003), Rigollet and Tsybakov (2007), Bunea, Tsybakov and Wegkamp (2007), Goldenshluger (2009);

- Exponential weights, Leung and Barron (2006), Dalalyan and Tsybakov (2008), Rigollet and Tsybakov (2011);
- Risk hull method, Cavalier and Golubev (2006);
- Blockwise Stein method, Cai (1999), Cavalier and Tsybakov (2001), Rigollet (2006).

Some of aforementioned papers deal with not only adaptation over the scale of functional classes but contain sharp oracle inequalities [about oracle approach and its relation to adaptive estimation; see, e.g., Goldenshluger and Lepski (2013) and the references therein]. Without any doubt, the adaptation under $\mathbb{L}_2$-loss is the best developed area of the adaptive estimation. A rather detailed overview and some new ideas related to this topic can be found in the recent paper Baraud, Giraud and Huet (2014).

The adaptive estimation under $\mathbb{L}_p$-loss, $1 \le p \le \infty$ was initiated by Lepskiĭ (1991) over the collection of Hölder classes, that is, $\mathbb{N}_{\infty,1}(\beta, L)$. The asymptotics of minimax risk is given by

$$\phi_\varepsilon\big(\mathbb{N}_{\infty,1}(\beta, L)\big) \sim \begin{cases} \varepsilon^{2\beta/(2\beta+1)}, & p \in [1,\infty); \\ \big(\varepsilon^2|\ln(\varepsilon)|\big)^{\beta/(2\beta+1)}, & p = \infty. \end{cases}$$

The author constructed the optimally-adaptive estimator which is obtained by the selection from the family of piecewise polynomial estimators. Selection rule is based on pairwise comparison of estimators (bias-majorant tradeoff). Some sharp results were obtain in Lepskiĭ (1992b), where an *efficient* adaptive estimator was proposed in the case of $\mathbb{L}_\infty$-loss; see also Tsybakov (1998).

Recent development in adaptive univariate density estimation under $\mathbb{L}_\infty$-loss can be found in Giné and Nickl (2009), Gach, Nickl and Spokoiny (2013). Another "extreme" case, the estimation under $\mathbb{L}_1$-loss, was scrutinized by Devroye and Lugosi (1996, 1997).

The consideration of the classes of inhomogeneous functions in nonparametric statistics was started in Nemirovskiy (1985), where the minimax rates of convergence were established and minimax estimators were constructed in the case of generalized Sobolev classes. The adaptive estimation problem over the scale of Besov classes $\mathbb{B}^\beta_{r,q}(L)$ was studied for the first time in Donoho et al. (1996) in the framework of the density model. We note that $\mathbb{B}^\beta_{r,\infty} = \mathbb{N}_{r,1}(\beta, L)$, and although $\mathbb{B}^\beta_{r,q} \supset \mathbb{N}_{r,1}(\beta, L)$ for any $q \ge 1$ [see Nikol'skiĭ (1977)], one has $\mathbb{B}^\beta_{r,q} \bowtie \mathbb{N}_{r,1}(\beta, L)$.

The same problem in the univariate model (1.1) was studied in Lepski, Mammen and Spokoiny (1997). The asymptotics of minimax risk is given by

$$\phi_\varepsilon\big(\mathbb{B}^\beta_{r,q}(L)\big) \sim \begin{cases} \varepsilon^{2\beta/(2\beta+1)}, & (2\beta+1)r > p; \\ \big(\varepsilon^2|\ln(\varepsilon)|\big)^{(\beta-1/r+1/p)/(2\beta-2/r+1)}, & (2\beta+1)r \le p. \end{cases}$$

The set of parameters satisfying $r(2\beta+1) > p$ is called in the literature *the dense zone* and the case $r(2\beta+1) \le p$ is referred to *the sparse zone*. As it was shown in Donoho et al. (1996) hard threshold wavelet estimator is *nearly* adaptive over the scale of Besov classes. The latter means that the maximal risk of the proposed estimator differs from $\phi(\mathbb{B}^{\beta}_{r,q}(L))$ by logarithmic factor on the dense zone and on the boundary $(2\beta+1)r = p$. A similar result was proved in Lepski, Mammen and Spokoiny (1997), but for a completely different estimation procedure: for the first time, a local bandwidth selection scheme was used for the estimation of an entire function. Moreover, the computations of the maximal risk of the proposed estimator on $\mathbb{B}^{\beta}_{r,q}(L)$ was made by integration of the local oracle inequality.

It is important to emphasize that both aforementioned results were proved under the additional assumption

$$1 - (\beta r)^{-1} + (\beta p)^{-1} > 0. \tag{1.4}$$

Independently, an approach similar to Lepski, Mammen and Spokoiny (1997) was proposed in Goldenshluger and Nemirovski (1997). The authors constructed *nearly* adaptive estimation over the scale of generalized Sobolev classes.

The optimally adaptive estimator over the scale of Besov classes was built by Juditsky (1997). The estimation procedure is the hard threshold wavelet construction with random thresholds whose choices are based on some modifications of the comparison scheme proposed in Lepskiĭ (1991). Several years later similar a result was obtained by Johnstone and Silverman (2005). This estimation method is again a hard threshold wavelet estimator but with empirical Bayes selection of thresholds. Both results were obtained under additional condition $\beta > 1/r$ which is slightly stronger than (1.4). *Efficient* adaptive estimator over the scale of Besov classes under $\mathbb{L}_2$-loss was constructed in Zhang (2005) by use of empirical Bayes thresholding.

We finish this section by mentioning the papers Juditsky and Lambert-Lacroix (2004) and Reynaud-Bouret, Rivoirard and Tuleau-Malot (2011), where very interesting phenomena related to the adaptive density estimation under $\mathbb{L}_p$-loss with unbounded support were observed, and the paper Goldenshluger (2009), where $\mathbb{L}_p$-aggregation of estimators was proposed.

Multivariate function estimation. Much less is known when adaptive estimation of multivariate function is considered. The principal difficulty is related to the fact that the methods developed in the univariate case cannot be directly generalized to the multivariate setting.

In a series of papers from the late 1970s, Ibragimov and Hasminskii studied the problem of minimax estimation over $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ under $\mathbb{L}_p$-losses in different statistical models; see Hasminskii and Ibragimov (1990) and references therein. Note, however, that these authors treated only the case $r_i = p, i = 1, \dots, d$, which allowed them to prove that standard linear estimators (kernel, local polynomial, etc.) are minimax. The optimally adaptive estimator corresponding to the latter case was constructed in Goldenshluger and Lepski (2011) in the density model on $\mathbb{R}^d$.

The estimation over *isotropic* Besov class $\mathbb{B}^{\vec{\beta}}_{\vec{r},q}(\vec{L})$ was studied in Delyon and Juditsky (1996), where the authors established the asymptotic of minimax risk under $\mathbb{L}_p$-loss and constructed minimax estimators. Here the isotropy means that $\vec{\beta} = (\boldsymbol{\beta}, \dots, \boldsymbol{\beta})$, $\vec{r} = (\mathbf{r}, \dots, \mathbf{r})$ and $\vec{L} = (\mathbf{L}, \dots, \mathbf{L})$. The asymptotics of minimax risk is given by

$$\phi_\varepsilon\big(\mathbb{B}^{\vec{\beta}}_{\vec{r},q}(\vec{L})\big) \sim \begin{cases} \varepsilon^{2\boldsymbol{\beta}/(2\boldsymbol{\beta}+d)}, & (2\boldsymbol{\beta}+d)r > dp; \\ \big(\varepsilon^2 |\ln(\varepsilon)|\big)^{(\boldsymbol{\beta}-d/\mathbf{r}+d/p)(2\boldsymbol{\beta}-2d/\mathbf{r}+d)}, & (2\boldsymbol{\beta}+d)\mathbf{r} \le dp. \end{cases}$$

Nearly adaptive with respect to $\mathbb{L}_p$-risk, $1 \le p < \infty$, an estimator over a collection of *isotropic* Nikol'skii classes $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \bowtie \mathbb{B}^{\vec{\beta}}_{\vec{r},q}(\vec{L})$ was built by Goldenshluger and Lepski (2008). The proposed procedure is based on the special algorithm of local bandwidth selection from the family of kernel estimators. The corresponding upper bound for maximal risks is proved under the additional assumption $\boldsymbol{\beta} > d/\mathbf{r}$.

Apparently, the first results on the minimax adaptive estimation over anisotropic Besov classes $\mathbb{B}^{\vec{\beta}}_{\vec{r},q}(\vec{L})$ was obtained in Neumann (2000) under $\mathbb{L}_2$-loss in the model (1.1). The author proposed the minimax and then nearly adaptive procedures based on the original hard threshold wavelet construction. This result was obtained under nonstandard and quite restrictive assumptions imposed on $\vec{\beta}$ and $\vec{r}$.

Bertin (2005) considered the problem of adaptive estimation over the scale of anisotropic Hölder classes, that is, $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ with $r_i = \infty$ for any $i = 1, \dots, d$ under $\mathbb{L}_\infty$-loss. The asymptotics of minimax risk is given here by

$$\phi_\varepsilon\big(\mathbb{N}_{\vec{\infty},d}(\vec{\beta}, \vec{L})\big) \sim \big(\varepsilon^2 |\ln(\varepsilon)|\big)^{\beta/(2\beta+1)},$$

where $1/\beta = 1/\beta_1 + \cdots + 1/\beta_d$. The construction of the optimally adaptive estimator is based on the selection rule from the family of kernel estimators developed in Lepski and Levit (1998).

Akakpo (2012) studied the problem of adaptive estimation over the scale of anisotropic Besov classes $\mathbb{B}^{\vec{\beta}}_{\vec{r},q}(\vec{L})$ under $\mathbb{L}_2$-loss in multivariate density models on the unit cube. The construction of the optimally-adaptive estimator is based on the model selection approach, and it uses sophisticated approximation bounds. Note, however, that all results are proved in the situation where coordinates of the vector $\vec{r}$ are the same ($r_i = r, i = 1, \dots, d$).

For the first time the minimax and minimax adaptive estimation over the scale of anisotropic classes $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ under $\mathbb{L}_p$-loss in the multivariate model (1.1) was studied in full generality in Kerkyacharian, Lepski and Picard (2001, 2008).

To describe the results obtained in this paper we will need the following notation used in the sequel as well. Set $\omega^{-1} = (\beta_1 r_1)^{-1} + \cdots + (\beta_d r_d)^{-1}$, and define for any $1 \le s \le \infty$

$$\tau(s) = 1 - 1/\omega + 1/(s\beta), \qquad \varkappa(s) = \omega(2 + 1/\beta) - s.$$

In Kerkyacharian, Lepski and Picard (2001) under assumption

$$\tau(\infty) > 0, \qquad \sum_{i=1}^{d}\left[1/(r_i\beta_i) - 1/(p\beta_i)\right]_+ < 2/p \tag{1.5}$$

(called by the authors *the dense zone*), the following asymptotics of minimax risk was found:

$$\phi_\varepsilon\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})\big) \sim \varepsilon^{\beta/(2\beta+1)}.$$

In Kerkyacharian, Lepski and Picard (2008) under assumptions

$$\tau(\infty) > 0, \qquad \varkappa(p) \le 0, \qquad \vec{r} \in [1,p]^d \tag{1.6}$$

(called by the authors *the sparse zone*) the following asymptotics of minimax risk was found:

$$\phi_\varepsilon\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})\big) \sim \big(\varepsilon^2\big|\ln(\varepsilon)\big|\big)^{\tau(p)/(2\tau(2))}.$$

The authors built a *nearly* adaptive estimator with respect to $\mathbb{L}_p$-risk, $1 \le p < \infty$. Its construction is based on the pointwise bandwidths selection rule which differs from that presented in Lepski and Levit (1998) as well as from the construction developed several years later in Goldenshluger and Lepski (2008, 2014). It is important to emphasize that the method developed in the present paper is in some sense a "global" version of the aforementioned procedure.

The existence of an optimally-adaptive estimator as well as the asymptotics of minimax risk in the case, where assumptions (1.5) and (1.6) are not fulfilled, remained an open problem. Note also that assumption (1.4), which appeared in the univariate case, can be rewritten as $\tau(p) > 0$. The minimax as well as adaptive estimation in the case $\tau(p) \le 0$ was not investigated. One can suppose that a uniformly consistent estimator on $\mathbb{N}_{r,1}(\beta, L)$ does not exist if $\tau(p) \le 0$ since $\tau(p) > 0$ is the sufficient condition for the compact embedding of the univariate Nikol'skii space into $\mathbb{L}_p$; see Nikol'skiĭ (1977).

The attempt to shed light on aforementioned problems was recently undertaken in Goldenshluger and Lepski (2014) in the framework of the density estimation on $\mathbb{R}^d$. The authors are interested in adaptive estimation under $\mathbb{L}_p$-loss, $p \in [1,\infty)$ over the collection of functional classes

$$\mathbb{F}_\vartheta = \mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},M) := \mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L}) \cap \{f : \|f\|_\infty \le M\}, \qquad \vartheta = (\vec{\beta},\vec{r},\vec{L},M).$$

Adapting the results obtained in the latter paper to the observation model (1.1), we first state that the asymptotics of the minimax risk satisfies

$$\phi_\varepsilon\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},M)\big) \gtrsim \mu_\varepsilon^{\nu},$$

where

$$\nu = \begin{cases} \dfrac{\beta}{2\beta+1}, & \varkappa(p) > 0; \\ \dfrac{\tau(p)}{2\tau(2)}, & \varkappa(p) \le 0, \tau(\infty) > 0; \\ \dfrac{\omega}{p}, & \varkappa(p) \le 0, \tau(\infty) \le 0; \end{cases}$$

$$\mu_\varepsilon = \begin{cases} \varepsilon^2, & \varkappa(p) > 0 \text{ or } \varkappa(p) \le 0, \tau(\infty) \le 0; \\ \varepsilon^2 |\ln(\varepsilon)|, & \varkappa(p) \le 0, \tau(\infty) > 0. \end{cases}$$

It is important to note that the obtained lower bound remains true if $p = \infty$, which implies in particular that under $\mathbb{L}_\infty$-loss, there is no uniformly consistent estimator on $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$ if $\tau(\infty) \le 0$ [note that $\varkappa(\infty) = -\infty$].

The authors proposed *nearly* adaptive estimator, that is, an estimator whose maximal risk is proportional to $(\varepsilon^2|\ln(\varepsilon)|)^\nu$, whatever the value of the nuisance parameter $\vartheta = (\vec{\beta}, \vec{r}, \vec{L}, M)$ and $p \in [1, \infty)$.

Thus the existence of optimally-adaptive estimators remains an open problem. Moreover, all discussed results are obtained under additional assumption that the underlying function is uniformly bounded. We will see that the situation changes completely if this condition does not hold. The optimally-adaptive estimator over the scale of anisotropic Nikol'skii classes under $\mathbb{L}_\infty$-loss was constructed in Lepski (2013) under assumption $\tau(\infty) > 0$. Since $\tau(\infty) > 0$ implies automatically that $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M) = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ for some M completely determined by $\vec{L}$, the investigation under $\mathbb{L}_\infty$-loss is finalized.

We would like to finish our short overview by mentioning works where the adaptation is studied not only with respect to the smoothness properties of the underlying function but also with respect to some structural assumptions imposed on the statistical model:

- Composite function structure, Horowitz and Mammen (2007), Juditsky, Lepski and Tsybakov (2009), Baraud and Birgé (2014);
- Multi-index structure (single-index, projection pursuit etc.), Hristache et al. (2001), Goldenshluger and Lepski (2009), Lepski and Serdyukova (2014);
- Multiple index model in density estimation, Samarov and Tsybakov (2007);
- Independence structure in density estimation, Lepski (2013).

The problems of adaptive estimation over the scale of functional classes defined on some manifolds were studied [Kerkyacharian, Thanh and Picard (2011), Kerkyacharian, Nickl and Picard (2012)].

1.4. *Objectives*. Considering the collection of functional classes

$$\mathbb{F}_\vartheta = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \qquad \vartheta = (\vec{\beta}, \vec{r}, \vec{L}),$$

we want to answer on the following questions:

(1) *What is the optimal decay of the minimax risk for any fixed value of the nuisance parameter ϑ and norm index $p \in [1, \infty]$?*

(2) *Do optimally-adaptive estimators always exist?*

To realize this program we propose first a new data-driven selection rule from the family of kernel estimators with *varying bandwidths* and establish for it so-called $\mathbb{L}_p$-norm oracle inequality. Then we use this inequality in order to prove the adaptivity properties of the proposed estimation procedure.

Let us discuss our approach more in detail. Throughout of the paper we will use the following notation. For any $u, v \in \mathbb{R}^d$ the operations and relations u/v, uv, $u \vee v$, $u \wedge v$, $u < v$, $au, a \in \mathbb{R}$, are understood in a coordinate-wise sense, and $|u|$ stands for Euclidean norm of u. All integrals are taken over $\mathbb{R}^d$ unless the domain of integration is specified explicitly. For any real a its positive part is denoted by $(a)_+$, and $\lfloor a \rfloor$ is used for its integer part.

Kernel estimator with varying bandwidth. Put $\mathfrak{H} = \{\mathfrak{h}_s = e^{-s-2}, s \in \mathbb{N}\}$, and denote by $\mathfrak{S}_1$ the set of all measurable functions defined on $(-b, b)^d$ and taking values in $\mathfrak{H}$. Introduce

$$\mathfrak{S}_d = \bigl\{\vec{h} : (-b, b)^d \to \mathfrak{H}^d : \vec{h}(x) = \bigl(h_1(x), \dots, h_d(x)\bigr),\\ x \in (-b, b)^d, h_i \in \mathfrak{S}_1, i = \overline{1, d}\bigr\}.$$

Let $K : \mathbb{R}^d \to \mathbb{R}$ be a function satisfying $\int K = 1$. With any $\vec{h} \in \mathfrak{S}_d$ we associate the function

$$K_{\vec{h}}(t, x) = V_{\vec{h}}^{-1}(x) K\biggl(\frac{t - x}{\vec{h}(x)}\biggr), \qquad t \in \mathbb{R}^d, x \in (-b, b)^d,$$

where $V_{\vec{h}}(x) = \prod_{i=1}^d h_i(x)$. Let $\mathfrak{S}^*$ be a given subset of $\mathfrak{S}_d$. Consider the family of estimators

$$\mathcal{F}(\mathfrak{S}^*) = \bigl\{\hat{f}_{\vec{h}}(x) = X_\varepsilon\bigl(K_{\vec{h}}(\cdot, x)\bigr), \vec{h} \in \mathfrak{S}^*, x \in (-b, b)^d\bigr\}. \tag{1.7}$$

We will call these estimators kernel estimators with varying bandwidth. This type of estimator was introduced in Müller and Stadtmüller (1987) in the context of cross-validation technique.

We will be particulary interested in the set $\mathfrak{S}^* = \mathfrak{S}_d^{\text{const}} \subset \mathfrak{S}_d$, which consists of constant functions. Note that if $\vec{h} \in \mathfrak{S}_d^{\text{const}}$ we come to the standard definition of kernel estimators in a white Gaussian noise model.

In view of (1.1) we have the following decomposition which will be useful in the sequel:

$$\begin{aligned} \hat{f}_{\vec{h}}(x) - f(x) &= \int K_{\vec{h}}(t, x)\bigl[f(t) - f(x)\bigr]\nu_d(\mathrm{d}t) + \varepsilon \xi_{\vec{h}}(x), \\ \xi_{\vec{h}}(x) &= \int K_{\vec{h}}(t, x) W(\mathrm{d}t). \end{aligned} \tag{1.8}$$

We note that $\xi_{\vec{h}}$ is a centered Gaussian random field on $(-b,b)^d$ with the covariance function

$$V_{\vec{h}}^{-1}(x)V_{\vec{h}}^{-1}(y)\int K\left(\frac{t-x}{\vec{h}(x)}\right)K\left(\frac{t-y}{\vec{h}(y)}\right)\nu_d(\mathrm{d}t), \qquad x,y\in(-b,b)^d.$$

Oracle approach. Our goal is to propose a data-driven (based on X^ε) selection procedure from the collection $\mathcal{F}(\mathfrak{S}^*)$ and establish for it $\mathbb{L}_p$-norm oracle inequality. More precisely we construct the random field $(\vec{\mathbf{h}}(x), x\in(-b,b)^d)$ completely determined by the observation X^ε, such that $x\mapsto\vec{\mathbf{h}}(x)$ belongs to $\mathfrak{S}^*$, and prove that for any $p\in[1,\infty]$, $q\geq 1$ and $\varepsilon>0$ small enough,

$$\mathcal{R}_\varepsilon^{(p)}[\hat{f}_{\vec{\mathbf{h}}};f]\leq\Upsilon_1\inf_{\vec{h}\in\mathfrak{S}^*}A_{p,q}^{(\varepsilon)}(f,\vec{h})+\Upsilon_2\varepsilon. \tag{1.9}$$

Here Υ_1 and Υ_2 are numerical constants depending on d,p,q,b and K only, and inequality (1.9) is established for any function $f\in\mathbb{L}_p(\mathbb{R}^d,\nu_d)\cap\mathbb{L}_2(\mathbb{R}^d,\nu_d)$. We call (1.9) an $\mathbb{L}_p$*-norm oracle inequality*.

We provide with explicit expression of the functional $A_{p,q}^{(\varepsilon)}(\cdot,\cdot)$ that allows us to derive different minimax adaptive results from the unique $\mathbb{L}_p$-norm oracle inequality. In this context it is interesting to note that in the "extreme cases" $p=1$ and $p=\infty$, it suffices to select the estimator from the family $\mathcal{F}(\mathfrak{S}_d^{\mathrm{const}})$. When $p\in(1,\infty)$, the oracle inequality (1.9) as well as the selection from the family $\mathcal{F}(\mathfrak{S}^*)$ will be done for some special choice of the bandwidth's set $\mathfrak{S}^*$. We will see that the restrictions imposed on $\mathfrak{S}^*$ are rather weak, which will allow us to prove very strong adaptive results presented in Section 3.

1.5. *Organization of the paper.* In Section 2 we present our selection rule and formulate for it $\mathbb{L}_p$-norm oracle inequality, Theorem 1. Its consequence related to the selection from the family $\mathfrak{S}_d^{\mathrm{const}}$ is established in Corollary 1. Section 3 is devoted to adaptive estimation over the collection of anisotropic Nikol'skii classes. The lower bound result is formulated in Theorem 2, and the adaptive upper bound is presented in Theorem 3. In Section 4 we discuss open problems in adaptive minimax estimation in different statistical models. Proofs of main results are given in Sections 5–7, and all technical lemmas are proven in the Appendix.

2. Selection rule and $\mathbb{L}_p$-norm oracle inequality.

2.1. *Functional classes of bandwidths.* Put for any $\vec{h}\in\mathfrak{S}_d$ and any $\mathbf{s}=(s_1,\dots,s_d)\in\mathbb{N}^d$

$$\Lambda_{\mathbf{s}}[\vec{h}]=\bigcap_{j=1}^d\Lambda_{s_j}[h_j], \qquad \Lambda_{s_j}[h_j]=\{x\in(-b,b)^d:h_j(x)=\mathfrak{h}_{s_j}\}.$$

Let $\varkappa \in (0,1)$ and $\mathfrak{L} > 0$ be given constants. Define

$$\mathbb{H}_d(\varkappa, \mathfrak{L}) = \Big\{\vec{h} \in \mathfrak{S}_d : \sum_{\mathbf{s}\in\mathbb{N}^d} \nu_d^{\varkappa}\big(\Lambda_{\mathbf{s}}[\vec{h}]\big) \le \mathfrak{L}\Big\}.$$

We remark that obviously $\mathfrak{S}_d^{\mathrm{const}} \subset \mathbb{H}_d(\varkappa, \mathfrak{L})$ for any $\varkappa \in (0,1)$ and $\mathfrak{L} = (2b)^{d\varkappa}$.

Put $\mathbb{N}_p^* = \{\lfloor p \rfloor + 1, \lfloor p \rfloor + 2, \dots\}$, and define for any $\mathcal{A} \ge e^d$

$$\mathbb{B}(\mathcal{A}) = \bigcup_{r\in\mathbb{N}_p^*} \mathbb{B}_r(\mathcal{A}), \qquad \mathbb{B}_r(\mathcal{A}) = \big\{\vec{h} \in \mathfrak{S}_d : \|V_{\vec{h}}^{-1/2}\|_{rp/(r-p)} \le \mathcal{A}\big\}.$$

Later on, in the case $p \in (1,\infty)$, we will be interested in selection from the family $\mathcal{F}(\mathbb{H})$, where $\mathbb{H}$ is an arbitrary subset of $\mathbb{H}_d(\varkappa, \mathfrak{L}, \mathcal{A}) := \mathbb{H}_d(\varkappa, \mathfrak{L}) \cap \mathbb{B}(\mathcal{A})$, $\varkappa \in (0, 1/d)$, with some special choice $\mathcal{A} = \mathcal{A}_\varepsilon \to \infty$, $\varepsilon \to 0$.

The following notation related to the functional class $\mathbb{B}(\mathcal{A})$ will be exploited in the sequel. For any $\vec{h} \in \mathbb{B}(\mathcal{A})$, define

$$\mathbb{N}_p^*(\vec{h}, \mathcal{A}) = \mathbb{N}_p^* \cap \big[r_{\mathcal{A}}(\vec{h}), \infty\big), \qquad r_{\mathcal{A}}(\vec{h}) = \inf\big\{r \in \mathbb{N}_p^* : \vec{h} \in \mathbb{B}_r(\mathcal{A})\big\}. \tag{2.1}$$

Obviously $r_{\mathcal{A}}(\vec{h}) < \infty$ for any $\vec{h} \in \mathbb{B}(\mathcal{A})$.

Assumptions imposed on the kernel K. Let $a \ge 1$ and $A > 0$ be fixed.

ASSUMPTION 1. There exists $\mathcal{K} : \mathbb{R} \to \mathbb{R}$ such that $\int \mathcal{K} = 1$, $\operatorname{supp}(K) \subset [-a, a]$ and:

$$\text{(i)} \qquad \big|\mathcal{K}(s) - \mathcal{K}(t)\big| \le A|s-t| \qquad \forall s, t \in \mathbb{R};$$

$$\text{(ii)} \qquad K(x) = \prod_{i=1}^{d} \mathcal{K}(x_i) \qquad \forall x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Throughout the paper we will consider only kernel estimators with K satisfying Assumption 1.

2.2. *Upper functions and the choice of parameters.* Put

$$\mathfrak{h}_\varepsilon := e^{-\sqrt{|\ln(\varepsilon)|}}, \qquad \mathcal{A}_\varepsilon := e^{\ln^2(\varepsilon)}, \tag{2.2}$$

and let $\mathfrak{S}_d(\mathfrak{h}_\varepsilon) \subset \mathfrak{S}_d$ consist of the functions $\vec{h}$, taking values in $\mathfrak{H}^d(\mathfrak{h}_\varepsilon) := \mathfrak{H}^d \cap (0, \mathfrak{h}_\varepsilon]^d$.

Set $C_2(r) = C_2(r, d\varkappa, (2\mathfrak{L})^d)$, and define for any $\vec{h} \in \mathbb{B}(\mathcal{A}_\varepsilon)$,

$$\widetilde{\Psi}_{\varepsilon,p}(\vec{h}) = C_1 \Big\|\sqrt{\big|\ln(\varepsilon V_{\vec{h}})\big|}\, V_{\vec{h}}^{-1/2}\Big\|_p, \qquad p \in [1,\infty],$$

$$\overline{\Psi}_{\varepsilon,p}(\vec{h}) = \inf_{r\in\mathbb{N}_p^*(\vec{h},\mathcal{A}_\varepsilon)} C_2(r) \big\|V_{\vec{h}}^{-1/2}\big\|_{rp/(r-p)}, \qquad p \in [1,\infty).$$

Introduce finally

$$\Psi_{\varepsilon,p}(\vec{h}) = \begin{cases} \widetilde{\Psi}_{\varepsilon,p}(\vec{h}) \wedge \overline{\Psi}_{\varepsilon,p}(\vec{h}), & \vec{h} \in \mathbb{B}(\mathcal{A}_\varepsilon) \cap \mathfrak{S}_d(\mathfrak{h}_\varepsilon),\ p \in [1,\infty); \\ \widetilde{\Psi}_{\varepsilon,p}(\vec{h}), & \vec{h} \in \mathbb{B}(\mathcal{A}_\varepsilon) \setminus \mathfrak{S}_d(\mathfrak{h}_\varepsilon),\ p \in [1,\infty]. \end{cases}$$

Some remarks are in order.

(1) The constant C_1 depends on $\mathcal{K}$, d, p and b, and its explicit expression is given in Section 5.1. The explicit expression of the quantity $C_2(r,\tau,\mathcal{L}), r > p, \tau \in (0,1), \mathcal{L} > 0$, can be found in Lepski (2015), Section 3.2.2. Its definition is rather involved, and since it will not be exploited in the sequel, we omit the definition of the latter quantity in the present paper. Here we only mention that $C_2(\cdot,\tau,\mathcal{L}) : (p,\infty) \to \mathbb{R}_+$ is bounded on each bounded interval. However, $C_2(r,\tau,\mathcal{L}) \to \infty, r \to \infty$.

(2) The selection rule presented below exploits heavily the fact that $\{\Psi_{\varepsilon,p}(\vec{h}), \vec{h} \in \mathbb{H}\}$, $p \in [1,\infty]$ is the upper function for the collection $\{\|\xi_{\vec{h}}\|_p, \vec{h} \in \mathbb{H}\}$. Here the random field $\xi_{\vec{h}}$ appeared in the decomposition (1.8) of the kernel estimator, and $\mathbb{H}$ is an arbitrary countable subset of $\mathbb{H}_d(\varkappa,\mathfrak{L},\mathcal{A}_\varepsilon)$. The latter result was recently proved in Lepski (2015), and it is presented in Proposition 1, Section 5.2 of the present paper.

(3) The choice of $\mathfrak{h}_\varepsilon$ and $\mathcal{A}_\varepsilon$ is mostly dictated by the following simple observation, which will be used for proving adaptive results presented in Section 3:

$$\lim_{\varepsilon\to 0} \varepsilon^{-a}\mathfrak{h}_\varepsilon = \infty, \qquad \lim_{\varepsilon\to 0} \varepsilon^{a}\mathcal{A}_\varepsilon = \infty \qquad \forall a > 0. \tag{2.3}$$

The general relation between parameters $\mathfrak{h}_\varepsilon$ and $\mathcal{A}_\varepsilon$ can be found in Lepski (2015).

2.3. *Selection rule.* Let $\mathbb{H}$ be a countable subset of $\mathbb{H}_d(\varkappa,\mathfrak{L},\mathcal{A}_\varepsilon)$. Define

$$\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) = \sup_{\vec{\eta}\in\mathbb{H}} \big[\|\hat{f}_{\vec{h}\vee\vec{\eta}} - \hat{f}_{\vec{\eta}}\|_p - \varepsilon\Psi_{\varepsilon,p}(\vec{h}\vee\vec{\eta}) - \varepsilon\Psi_{\varepsilon,p}(\vec{\eta})\big]_+, \qquad \vec{h} \in \mathbb{H}. \tag{2.4}$$

Our selection rule is given now by $\vec{\mathbf{h}}_0 = \arg\inf_{\vec{h}\in\mathbb{H}}\{\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\}$. Since $\vec{\mathbf{h}}_0$ does not necessarily belong to $\mathbb{H}$, we define finally $\vec{\mathbf{h}} \in \mathbb{H}$ from the relation

$$\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{\mathbf{h}}) + \varepsilon\Psi_{\varepsilon,p}(\vec{\mathbf{h}}) \le \widehat{\mathcal{R}}_{\mathbb{H}}(\vec{\mathbf{h}}_0) + \varepsilon\Psi_{\varepsilon,p}(\vec{\mathbf{h}}_0) + \varepsilon, \tag{2.5}$$

which leads to the estimator $\hat{f}_{\vec{\mathbf{h}}}$.

REMARK 1. We restrict ourselves by consideration of countable subsets of $\mathbb{H}_d(\varkappa,\mathfrak{L},\mathcal{A}_\varepsilon)$ in order not to discuss the measurability of $\hat{f}_{\vec{\mathbf{h}}}$. Formally, the proposed selection rule can be applied for any $\mathbb{H} \subseteq \mathbb{H}_d(\varkappa,\mathfrak{L},\mathcal{A}_\varepsilon)$ for which final estimator can be correctly defined.

2.4. $\mathbb{L}_p$*-norm oracle inequality.* For any $\vec{\mathrm{h}} \in \mathfrak{S}_d$ define

$$S_{\vec{\mathrm{h}}}(x, f) = \int_{\mathbb{R}^d} K_{\vec{\mathrm{h}}}(t - x) f(t) \nu_d(\mathrm{d}t), \qquad x \in \mathbb{R}^d,$$

which is understood as kernel approximation (smoother) of the function f at a point x.

For any $\vec{h}, \vec{\eta} \in \mathfrak{S}_d$ introduce also

$$(2.6) \qquad \begin{aligned} B_{\vec{h},\vec{\eta}}(x, f) &:= \left|S_{\vec{h}\vee\vec{\eta}}(x, f) - S_{\vec{\eta}}(x, f)\right|, \\ B_{\vec{h}}(x, f) &= \left|S_{\vec{h}}(x, f) - f(x)\right|, \end{aligned}$$

and define finally for any $p \in [1, \infty]$

$$(2.7) \qquad \mathcal{B}^{(p)}_{\vec{h}}(f) = \sup_{\vec{\eta}\in\mathbb{H}} \left\| B_{\vec{h},\vec{\eta}}(\cdot, f)\right\|_p + \left\| B_{\vec{h}}(\cdot, f)\right\|_p.$$

THEOREM 1. *Let Assumption 1 be fulfilled, and let* $p \in [1, \infty]$, $q \geq 1$, $\varkappa \in (0, 1/d)$ *and* $\mathfrak{L} \geq 1$ *be fixed. Then there exists* $\varepsilon(q) > 0$ *such that for any* $\varepsilon \leq \varepsilon(q)$ *and* $\mathbb{H} \subseteq \mathbb{H}_d(\varkappa, \mathfrak{L}, \mathcal{A}_\varepsilon)$,

$$\mathcal{R}^{(p)}_\varepsilon[\hat{f}_{\vec{\mathbf{h}}}; f] \leq 5 \inf_{\vec{h}\in\mathbb{H}} \left\{\mathcal{B}^{(p)}_{\vec{h}}(f) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\right\} + 9(C_3 + C_4 + 2)\varepsilon \qquad \forall f \in \mathbb{L}_p\left(\mathbb{R}^d, \nu_d\right) \cap \mathbb{L}_2\left(\mathbb{R}^d, \nu_d\right).$$

The quantities C_3 and C_4 depend on $\mathcal{K}, p, q, b$ and d only, and their explicit expressions are presented in the Section 5.1.

Some consequences. Selection rule (2.5) deals with the family of kernel estimators with varying bandwidths. This allows, in particular, to apply $\mathbb{L}_p$-norm oracle inequality established in Theorem 1 to adaptive estimation over the collection of inhomogeneous and anisotropic functional classes. However, in some cases it suffices to select from much less "massive" set of bandwidths, namely from $\mathfrak{S}^{\mathrm{const}}_d$. In this case one can speak about standard multi-bandwidth selection. In particular, in the next section we will show that the selection from $\mathfrak{S}^{\mathrm{const}}_d$ leads to an optimally adaptive estimator over anisotropic Nikol'skii classes if $p = \{1, \infty\}$. Moreover, considering $\mathfrak{S}^{\mathrm{const}}_d$, we simplify considerably the "approximation error" $\mathcal{B}^{(p)}_{\vec{h}}(f)$ as well as the upper function $\Psi_{\varepsilon,p}(\cdot)$. The following corollary of Theorem 1 will be proved in Section 5.2.

Set $C_{2,p} = (2b)^{d/p} \inf_{r\in\mathbb{N}^*_p} C_2(r)$, and define for any $\vec{h} \in \mathfrak{S}^{\mathrm{const}}_d(\mathfrak{h}_\varepsilon) := \mathfrak{S}^{\mathrm{const}}_d \cap \mathfrak{S}_d(\mathfrak{h}_\varepsilon)$,

$$(2.8) \qquad \begin{aligned} \Psi^{(\mathrm{const})}_{\varepsilon,p}(\vec{h}) &= C_{2,p} V_{\vec{h}}^{-1/2}, \qquad p \in [1, \infty), \\ \Psi^{(\mathrm{const})}_{\varepsilon,\infty}(\vec{h}) &= C_1 \sqrt{\left|\ln(\varepsilon V_{\vec{h}})\right|}\, V_{\vec{h}}^{-1/2}. \end{aligned}$$

Let $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ be the canonical basis in $\mathbb{R}^d$. For any $\vec{h} \in \mathfrak{S}_d^{\mathrm{const}}$ introduce

$$b_{\vec{h},j}(x) = \sup_{s:\mathfrak{h}_s \le h_j} \left| \int_{\mathbb{R}} \mathcal{K}(u) f(x + u\mathfrak{h}_s \mathbf{e}_j)\nu_1(\mathrm{d}u) - f(x) \right|, \tag{2.9}$$
$$j = 1, \dots, d.$$

Define finally $\mathbb{H}_\varepsilon^{\mathrm{const}} = \mathfrak{S}_d^{\mathrm{const}}(\mathfrak{h}_\varepsilon) \cap \{\vec{h} : V_{\vec{h}} \ge (2b)^{d/p}\mathcal{A}_\varepsilon^{-2}\}$, and let $\hat{f}_{\vec{\mathbf{h}}}^{(\mathrm{const})}$ be the estimator obtained by selection rule (2.5) with $\mathbb{H} = \mathbb{H}_\varepsilon^{\mathrm{const}}$ and $\Psi_{\varepsilon,p}(\vec{h})$ replaced by $\Psi_{\varepsilon,p}^{(\mathrm{const})}(\vec{h})$ given in (2.8).

COROLLARY 1. *Let Assumption 1 be fulfilled, and let $p \in [1, \infty]$ and $q \ge 1$ be fixed. Then there exists $\varepsilon(q) > 0$ such that for any $\varepsilon \le \varepsilon(q)$, $\mathbb{H} \subseteq \mathbb{H}_\varepsilon^{\mathrm{const}}$ and $f \in \mathbb{L}_p(\mathbb{R}^d, \nu_d) \cap \mathbb{L}_2(\mathbb{R}^d, \nu_d)$*

$$\mathcal{R}_\varepsilon^{(p)}[\hat{f}_{\vec{\mathbf{h}}}^{(\mathrm{const})}; f] \le 5 \inf_{\vec{h}\in\mathbb{H}} \left\{ 3\|\mathcal{K}\|_{1,\mathbb{R}}^d \sum_{j=1}^d \|b_{\vec{h},j}\|_p + \varepsilon\Psi_{\varepsilon,p}^{(\mathrm{const})}(\vec{h}) \right\}$$
$$+ 9(C_3 + C_4 + 2)\varepsilon.$$

We remark that since $\mathbb{H}_\varepsilon^{\mathrm{const}}$ is finite a selected multi-bandwidth, $\vec{\mathbf{h}} \in \mathbb{H}$ is given by

$$\vec{\mathbf{h}} = \arg \inf_{\vec{h}\in\mathbb{H}} \{\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) + \varepsilon\Psi_{\varepsilon,p}^{(\mathrm{const})}(\vec{h})\}.$$

3. Adaptive estimation. In this section we study properties of the estimator defined in Section 2.3. The $\mathbb{L}_p$-norm oracle inequalities obtained Theorem 1 and Corollary 1 can be viewed as the initial step in bounding the $\mathbb{L}_p$-risk of this estimator on the anisotropic Nikol'skii classes.

3.1. *Anisotropic Nikol'skii classes*. Recall that $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ denotes the canonical basis of $\mathbb{R}^d$. For the function $g : \mathbb{R}^d \to \mathbb{R}^1$ and real number $u \in \mathbb{R}$, define *the first order difference operator with step size u in direction of the variable x_j* by

$$\Delta_{u,j} g(x) = g(x + u\mathbf{e}_j) - g(x), \qquad j = 1, \dots, d.$$

By induction, the kth order difference operator with step size u in direction of the variable x_j is defined as

$$\Delta_{u,j}^k g(x) = \Delta_{u,j}\Delta_{u,j}^{k-1} g(x) = \sum_{l=1}^k (-1)^{l+k} \binom{k}{l} \Delta_{ul,j} g(x). \tag{3.1}$$

DEFINITION 1. For given vectors $\vec{r} = (r_1, \dots, r_d)$, $r_j \in [1, \infty]$, $\vec{\beta} = (\beta_1, \dots, \beta_d)$, $\beta_j > 0$, and $\vec{L} = (L_1, \dots, L_d)$, $L_j > 0$, $j = 1, \dots, d$, we say that function $g : \mathbb{R}^d \to \mathbb{R}^1$ belongs to the anisotropic Nikol'skii class $\bar{\mathbb{N}}_{\vec{r},d}(\vec{\beta}, \vec{L})$ if:

(i) $\|g\|_{r_j,\mathbb{R}^d} \le L_j$ for all $j=1,\dots,d$;

(ii) for every $j=1,\dots,d$ there exists natural number $k_j > \beta_j$ such that

$$\big\|\Delta_{u,j}^{k_j} g\big\|_{r_j,\mathbb{R}^d} \le L_j |u|^{\beta_j} \qquad \forall u \in \mathbb{R}, \forall j=1,\dots,d. \tag{3.2}$$

Recall that the consideration of white Gaussian noise model requires $f \in \mathbb{L}_2(\mathbb{R}^d)$ that is not always guaranteed by $f \in \bar{\mathbb{N}}_{\vec{r},d}(\vec{\beta},\vec{L})$. So, later on we will study the functional classes $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L}) = \bar{\mathbb{N}}_{\vec{r},d}(\vec{\beta},\vec{L}) \cap \mathbb{L}_2(\mathbb{R}^d)$, which we will also call anisotropic Nikol'skii classes. Some conditions with guaranteed $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L}) = \bar{\mathbb{N}}_{\vec{r},d}(\vec{\beta},\vec{L})$ can be found in Section 7.1.

3.2. *Main results.* Let $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})$ be the anisotropic Nikol'skii functional class. Put

$$\frac{1}{\beta} := \sum_{j=1}^{d} \frac{1}{\beta_j}, \qquad \frac{1}{\omega} := \sum_{j=1}^{d} \frac{1}{\beta_j r_j}, \qquad L_\beta := \prod_{j=1}^{d} L_j^{1/\beta_j},$$

and define for any $1 \le s \le \infty$,

$$\tau(s) = 1 - 1/\omega + 1/(s\beta), \qquad \varkappa(s) = \omega(2+1/\beta) - s.$$

The following obvious relation will be useful in the sequel:

$$\frac{\varkappa(s)}{\omega s} = \frac{2-s}{s} + \tau(s). \tag{3.3}$$

Set finally $p^* = [\max_{j=1,\dots,d} r_l] \vee p$, and introduce

$$\mathfrak{a} = \begin{cases} \dfrac{\beta}{2\beta+1}, & \varkappa(p) > 0; \\ \dfrac{\tau(p)}{2\tau(2)}, & \varkappa(p) \le 0, \tau(p^*) > 0; \\ \dfrac{\omega(p^*-p)}{p(p^*-\omega(2+1/\beta))}, & \varkappa(p) \le 0, \tau(p^*) \le 0, p^* > p; \\ 0, & \varkappa(p) \le 0, \tau(p^*) \le 0; p^* = p. \end{cases}$$

$$\delta_\varepsilon = \begin{cases} L_\beta \varepsilon^2, & \varkappa(p) > 0; \\ L_\beta \varepsilon^2 |\ln(\varepsilon)|, & \varkappa(p) \le 0, \tau(p^*) \le 0; \\ L_\beta^{(1-2/p)/\tau(p)} \varepsilon^2 |\ln(\varepsilon)|, & \varkappa(p) \le 0, \tau(p^*) > 0. \end{cases}$$

3.2.1. *Lower bound of minimax risk.*

THEOREM 2. *Let $q \ge 1$, $L_0 > 0$ and $1 \le p \le \infty$ be fixed. Then for any $\vec{\beta} \in (0,\infty)^d$, $\vec{r} \in [1,\infty]^d$ and $\vec{L} \in [L_0,\infty)^d$, there exists $c > 0$ independent of $\vec{L}$ such* *that*

$$\liminf_{\varepsilon\to 0}\inf_{\tilde{f}_\varepsilon}\sup_{f\in\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})}\delta_\varepsilon^{-\mathfrak{a}}\mathcal{R}_\varepsilon^{(p)}[\tilde{f}_\varepsilon; f]\geq c,$$

where infimum is taken over all possible estimators.

Let us make several remarks.

1^0. *Case* $p^*=p$. Taking into account (3.3) we note that there is no uniformly consistent estimator over $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})$ if

$$\tau(p)1_{[2,\infty)}(p)+\varkappa(p)1_{[1,2)}(p)\leq 0, \tag{3.4}$$

and this result seems to be new. As it will follow from the next theorem the latter condition is necessary and sufficient for nonexistence of uniformly consistent estimators over $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})$ under $\mathbb{L}_p$-loss, $1\leq p\leq\infty$. In the case of $\mathbb{L}_\infty$-loss, (3.4) is reduced to $\omega\leq 1$, and a similar result was recently proved in Goldenshluger and Lepski (2014) for the density model.

It is worth noting that the assumption $\tau(p)>0$ in the considered case $p=p^*$ is the sufficient condition for the compact embedding of the corresponding Nikol'skii class to another class which is a compact in $\mathbb{L}_p(\mathbb{R}^d)$. Recall that there exists the compliance between the existence of a uniform consistence estimator and the compactness of the class on which the maximal risk is considered. In this context, the discussed result of the theorem perhaps implies that $\tau(p)>0$, $p\geq 2$ is also the necessary condition for the latter embedding. However, this problem lies beyond the scope of the present paper.

2^0. *Case* $\varkappa(p)\leq 0, \tau(p^*)\leq 0, p^*>p$. The lower bound for minimax risk, given in this case by

$$\big(L_\beta\varepsilon^2\big|\ln(\varepsilon)\big|\big)^{(\omega(p^*-p))/(p(p^*-\omega(2+1/\beta))},$$

is new. It is interesting that the latter case does not appear in dimension 1 or, more generally, when isotropic Nikol'skii classes are considered. Indeed, if $r_l=r$ for all $l=1,\dots,d$, then $p^*>p$ means $r>p$ that, in its turn, implies $\tau(p^*)=\tau(r)=1>0$. It is worth mentioning that we improve in order the lower bound, recently found in Goldenshluger and Lepski (2014), which corresponds formally to our case $p^*=\infty$.

3^0. *Case* $\varkappa(p)\leq 0, \tau(p^*)>0$. For the first time the same result was proved in Kerkyacharian, Lepski and Picard (2008) but under more restrictive assumption $\varkappa(p)\leq 0, \tau(\infty)>0$. Moreover, the dependence of the asymptotics of the minimax risk on $\vec{L}$ was not optimal.

4^0. *Case* $\varkappa(p)>0$. The presented lower bound of minimax risk became the statistical folklore since it is the minimax rate of convergence over anisotropic Hölder class ($r_l=\infty, l=1,\dots,d$). If so, the required result can be easily deduced from the embedding of a Hölder class to $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})$, whatever the value of $\vec{r}$.

However, the author was unable to find exact references and derived the announced result from the general construction used in the proof of Theorem 2. Moreover we are interested in finding not only the optimal decay of the minimax risk with respect to $\varepsilon \to 0$, but also its correct dependence on the radii $\vec{L}$.

3.2.2. *Upper bound for minimax risk. Optimally-adaptive estimator.* The results of this section will be derived from $\mathbb{L}_p$-norm oracle inequalities proved in Theorem 1 and Corollary 1.

Construction of kernel K. We will use the following specific kernel K [see, e.g., Kerkyacharian, Lepski and Picard (2001) or Goldenshluger and Lepski (2011)] in the definition of the estimator's family (1.7).

Let ℓ be an integer number, and let $w : [-1/(2\ell), 1/(2\ell)] \to \mathbb{R}^1$ be a function satisfying $\int w(y)\,\mathrm{d}y = 1$ and $w \in \mathbb{C}^1(\mathbb{R}^1)$. Put

$$w_\ell(y) = \sum_{i=1}^{\ell} \binom{\ell}{i} (-1)^{i+1} \frac{1}{i} w\left(\frac{y}{i}\right),$$
$$K(t) = \prod_{j=1}^{d} w_\ell(t_j), \qquad t = (t_1, \dots, t_d). \tag{3.5}$$

Set of bandwidths. Set $t_{k,n} = -(b+1) + (b+1)k2^{1-n}$, $k = 0, \dots, 2^n$, $n \in \mathbb{N}^*$, and let $\Delta_{k,n} = [t_{k,n}, t_{k+1,n})$, $k = 0, \dots, 2^{n-1}$, $\Delta_{k,n} = (t_{k,n}, t_{k+1,n}]$, $k = 2^{n-1} + 1, \dots, 2^n - 1$. Thus $\{\Delta_{k,n}, k = 0, \dots, 2^n\}$ forms the partition of $(-b-1, b+1)$ whatever $n \in \mathbb{N}^*$.

For any $n \in \mathbb{N}^*$, set also $\mathfrak{K}_n = \{0, \dots, 2^n\}^d$, and define

$$\Gamma_d(n) = \{\Delta^{(d)}_{\mathbf{k},n} = \Delta_{k_1,n} \times \cdots \times \Delta_{k_d,n}, \mathbf{k} = (k_1, \dots, k_d) \in \mathfrak{K}_n\}. \tag{3.6}$$

For any $n \in \mathbb{N}^*$ the collection of cubs $\Gamma_d(n)$ determines the partition of $(-b-1, b+1)^d$.

Denote by $\mathfrak{S}^{(n)}_d$, $n \in \mathbb{N}^*$ the set of all step functions defined on $(-b, b)^d$ with the steps belonging to $\Gamma_d(n) \cap (-b, b)^d$ and taking values in $\mathfrak{H}^d$.

Introduce finally for any $R > 0$,

$$\mathbb{H}_\varepsilon(R) = \mathbb{H}_d\big(1/(2d), R, \mathcal{A}_\varepsilon\big) \cap \left\{ \bigcup_{n \in \mathbb{N}^*} \mathfrak{S}^{(n)}_d \right\},$$

where $\mathcal{A}_\varepsilon$ is given in (2.2).

Let $\hat{f}^{(R)}_{\vec{\mathbf{h}}}$, $R > 0$ denote the estimator obtained by the selection rule (2.4)–(2.5) from the family of kernel estimators $\mathcal{F}(\mathbb{H}_\varepsilon(R))$ and $\hat{f}^{(\mathrm{const})}_{\vec{\mathbf{h}}}$ denote the estimator constructed in Corollary 1. Both constructions are made with the kernel K satisfying (3.5).

Adaptive upper bound. For any $\ell \in \mathbb{N}^*$ and $L_0 > 0$, set $\Theta = (0, \ell]^d \times [1, \infty]^d \times [L_0, \infty)^d$, and later on we will use the notation $\vartheta \in \Theta$ for the triplet $(\vec{\beta}, \vec{r}, \vec{L})$. Denote $\mathcal{P} = \Theta \times [1, \infty]$, and introduce

$$\mathcal{P}^{\text{consist}} = \{(\vartheta, p) \in \mathcal{P} : \tau(p)1_{[2,\infty)}(p) + \varkappa(p)1_{[1,2)}(p) > 0\} \\ \cup \{(\vartheta, p) \in \mathcal{P} : p^* > p\}.$$

The latter set consists of the class parameters and norm indexes for which a uniform consistent estimation is possible.

Let $V_p(\vec{L})$ be the quantity whose presentation is postponed to Section 7.4 since its expression is rather cumbersome. Put $L^* = \min_{j : r_j = p^*} L_j$, and introduce

$$\boldsymbol{\delta}_\varepsilon = \begin{cases} L_\beta \varepsilon^2, & \varkappa(p) \geq 0; \\ L_\beta (L^*)^{1/\mathfrak{a}} \varepsilon^2 |\ln(\varepsilon)|, & \varkappa(p) \leq 0, \tau(p^*) \leq 0; \\ V_p(\vec{L}) \varepsilon^2 |\ln(\varepsilon)|, & \varkappa(p) \leq 0, \tau(p^*) > 0. \end{cases}$$

THEOREM 3. *Let $q \geq 1$, $L_0 > 0$ and $\ell \in \mathbb{N}^*$ be fixed, and let $R = 3 + \sqrt{2b}$:*

(1) *For any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ such that $p \in (1, \infty)$, $\vec{r} \in (1, \infty]^d$ and $\varkappa(p) \neq 0$, there exists $C > 0$ independent of $\vec{L}$ for which*

$$\limsup_{\varepsilon \to 0} \sup_{f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})} \boldsymbol{\delta}_\varepsilon^{-\mathfrak{a}} \mathcal{R}_\varepsilon^{(p)} [\hat{f}_{\vec{\mathbf{h}}}^{(R)}; f] \leq C.$$

(2) *For any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$, $p \in \{1, \infty\}$, there exists $C > 0$ independent of $\vec{L}$ for which*

$$\limsup_{\varepsilon \to 0} \sup_{f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})} \boldsymbol{\delta}_\varepsilon^{-\mathfrak{a}} \mathcal{R}_\varepsilon^{(p)} [\hat{f}_{\vec{\mathbf{h}}}^{(\text{const})}; f] \leq C.$$

(3) *For any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ such that $p \in (1, \infty)$, $\vec{r} \in (1, \infty]^d$ and $\varkappa(p) = 0$, there exists $C > 0$ independent of $\vec{L}$ for which*

$$\limsup_{\varepsilon \to 0} \sup_{f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})} \boldsymbol{\delta}_\varepsilon^{-\mathfrak{a}} (|\ln(\varepsilon)|)^{1/p} \mathcal{R}_\varepsilon^{(p)} [\hat{f}_{\vec{\mathbf{h}}}^{(R)}; f] \leq C.$$

Some remarks are in order.

1^0. Combining the results of Theorems 2 and 3, we conclude that optimally-adaptive estimators under $\mathbb{L}_p$-loss exist over all parameter sets $\mathcal{P}^{\text{consist}}$ if $p \in \{1, \infty\}$. If $p \in (1, \infty)$, such estimators exist as well, except the boundary cases $\varkappa(p) = 0$ and $\min_{j=1,\dots,d} r_j = 1$. Note also that the adaptation is with respect to the parameter $\vec{\beta}$ is restricted on $(0, \ell]^d$. Hence we want to fix this parameter as large as possible but it should be done a priori since it is involved in our procedure.

2^0. We remark that the upper and lower bounds for minimax risk differ each other on the boundary $\varkappa(p) = 0$ only by $(|\ln(\varepsilon)|)^{1/p}$-factor. Using $(1, 1)$-weak type inequality for strong maximal operator [de Guzmán (1975)], one can prove adaptive upper bound on the boundary $\min_{j=1,\dots,d} r_j = 1$ containing an additional $(|\ln(\varepsilon)|)^{(d-1)/p}$-factor. Note, nevertheless, that exact asymptotics of minimax risk on both boundaries remains an open problem.

3^0. We obtain full classification of minimax rates over anisotropic Nikol'skii classes if $p \in \{1, \infty\}$ and "almost" full one (except the boundaries mentioned above) if $p \in (1, \infty)$. We can assert that $\delta_\varepsilon^{\mathfrak{a}}$ is minimax rate of convergence on $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ for any $\vec{\beta} \in (0, \infty)^d, \vec{r} \in (1, \infty]^d$ and $\vec{L} \in (0, \infty)^d$. Indeed, for given $\vec{\beta}$ and $\vec{L}$ one can choose $L_0 = \min_{j=1,\dots,d} L_j$ and the number ℓ, used in the kernel construction (3.5), as an any integer strictly larger than $\max_{j=1,\dots,d} \beta_j$.

4^0. We remark that the dependence of minimax rate on $\vec{L}$ is correct $(\delta_\varepsilon = \boldsymbol{\delta}_\varepsilon)$ if $\varkappa(p) \ge 0$. In spite of the cumbersome expression of the quantity $V_p(\vec{L})$, one can easily check that

$$V_p(\vec{L}) = L_\beta^{(1-2/p)/(\tau(p))}$$

if $L_j = L$ for any $j = 1, \dots, d$. Hence, under this restriction, $\delta_\varepsilon = \boldsymbol{\delta}_\varepsilon$ if $\varkappa(p) \le 0, \tau(p^*) > 0$ as well.

4. Open problems in adaptive estimation. The goal of this section is to discuss the directions in which adaptive multivariate function estimation will be developed. We do not pretend here to cover the whole spectrum of existing problems and mostly restrict ourselves by consideration of the adaptation over the scale of anisotropic classes. Moreover, we will concentrate on principal difficulties and the mathematical aspect of the problem, and we will not pay much attention to the technical details and practical applications. Although we will speak about adaptive estimation, it is important to realize that for the majority of problems discussed below, very little is known about the minimax approach.

4.1. *Abstract statistical model.* Let $(\mathcal{Y}^{(n)}, \mathfrak{A}^{(n)}, \mathbb{P}_f^{(n)}, f \in \mathbb{F})$ be the sequence of statistical experiments generated by observation $Y^{(n)}, n \in \mathbb{N}^*$. Let Λ be a set and $\rho: \Lambda \times \Lambda \to \mathbb{R}_+$ be a loss functional. The goal is to estimate the mapping $G: \mathbb{F} \to \Lambda$, and as an estimator we understand an $Y^{(n)}$-measurable Λ-valued map.

The quality of an estimation procedure $\tilde{G}_n$ on $\mathbb{F}$ is measured by the maximal risk

$$\mathcal{R}_n[\tilde{G}_n; \mathbb{F}] = \Big\{\sup_{f \in \mathbb{F}} \mathbb{E}_f^{(n)} \rho^q\big(\tilde{G}_n, G(f)\big)\Big\}^{1/q}, \qquad q \ge 1,$$

and as previously, $\phi_n(\mathbb{F}) = \inf_{\tilde{G}_n} \mathcal{R}_n[\tilde{G}_n; \mathbb{F}]$ denotes the minimax risk.

Assume that $\mathbb{F} \supset \bigcup_{\vartheta \in \Theta} \mathbb{F}_\vartheta$, where $\{\mathbb{F}_\vartheta, \vartheta \in \Theta\}$ is a given collection of sets.

PROBLEM I (Fundamental). Find necessary and sufficient conditions of existence of optimally-adaptive estimators, that is, the existence of an estimator $\hat{G}_n$

satisfying

$$\mathcal{R}_n[\hat{G}_n; \mathbb{F}_\vartheta] \sim \phi_n(\mathbb{F}_\vartheta) \qquad \forall \vartheta \in \Theta.$$

It is well known that optimally-adaptive estimators do not always exist; see Lepskiĭ (1990, 1992a), Efromovich and Low (1994), Cai and Low (2005). Hence the goal is to understand how the answer to the aforementioned question depends on the statistical model, underlying estimation problem (mapping G) and the collection of considered classes. The attempt to provide such classification was undertaken in Lepskiĭ (1992a), but we found there that sufficient conditions of existence as well as of nonexistence of optimally-adaptive estimators are too restrictive.

It is important to realize that the answers to the formulated problem may be different, even if the statistical model and the collection of functional classes are the same, and estimation problems are similar in nature. Indeed, consider univariate model (1.1), and let $\mathbb{F}_\vartheta = \mathbb{N}_{\infty,1}(\beta, L)$, $\vartheta = (\beta, L)$, be the collection of Hölder classes. Set

$$G_\infty(f) = \|f\|_\infty, \qquad G_2(f) = \|f\|_2.$$

As we know the optimally-adaptive estimator of f, say $\hat{f}_n$, under $\mathbb{L}_\infty$-loss was constructed in Lepskiĭ (1991). Moreover the asymptotics of minimax risk under $\mathbb{L}_\infty$-loss on $\mathbb{N}_{\infty,1}(\beta, L)$ coincides with asymptotics corresponding to the estimation of $G_\infty(\cdot)$. Therefore, $\hat{G}_n := G_\infty(\hat{f}_n)$ is an optimally-adaptive estimator for $G_\infty(\cdot)$. On the other hand, there is no optimally-adaptive estimator for $G_2(\cdot)$; see Cai and Low (2006).

4.2. *White Gaussian noise model.* Let us return to the problems studied in the present paper. Looking at the optimally-adaptive estimator proposed in Theorem 3, we conclude that its construction is not feasible. Indeed, it is based on the selection from very huge set of parameters, sometimes even infinite.

PROBLEM II (Feasible estimator). Find an optimally-adaptive estimator whose construction would be computationally reasonable.

At first glance, the interest in this problem is not related to the "practical applications" since the pointwise bandwidths selection rule from Goldenshluger and Lepski (2014) will do the job, although it is not theoretically optimal. We think that a "feasible solution" could bring new ideas and approaches to the construction of estimation procedures. One very interesting remark (prompted to the author by one of the referees) is that a solution being simultaneously statistically optimal and computationally feasible may not exist. In this context the negative answer on Problem II would be of big interest as well.

Another source of problems is structural adaptation. Let us consider one of the possible directions. Denote by $\mathcal{E}$ the set of all $d \times s$ real matrices, $1 \le s < d$. Introduce the following collection of functional classes:

$$\mathbb{F}_\vartheta = \mathbb{S}_{\vec{r},d}(\vec{\beta}, \vec{L}, E) := \{f : \mathbb{R}^d \to \mathbb{R} : f(x) = g(Ex),\, g \in \mathbb{N}_{\vec{r},p}(\vec{\beta}, \vec{L}),\, E \in \mathcal{E}\},$$
$$\vartheta = (\vec{\beta}, \vec{r}, \vec{L}, E).$$

PROBLEM III (Structural adaptation). Prove or disprove the existence of optimally-adaptive estimators over the collection $\mathbb{S}_{\vec{r},d}(\vec{\beta}, \vec{L}, E)$ under $\mathbb{L}_p$-loss.

Note that if $\vec{r} = (\infty, \dots, \infty)$ (Hölder case), an optimally adaptive estimator was constructed in Goldenshluger and Lepski (2009). A *nearly* adaptive estimator in the case $s = 1$ (single index constraint) and $d = 2$ was proposed in Lepski and Serdyukova (2014). Many other structural models like additive, projection pursuit or their generalization [see Goldenshluger and Lepski (2009)] can be studied as well.

4.3. *Density model.* Let $X_i, i = 1, \dots, n$, be *i.i.d.* d-dimensional random vectors with common probability density f. The goal is to estimate f under $\mathbb{L}_p$-loss on $\mathbb{R}^d$.

PROBLEM IV. Prove or disprove the existence of optimally-adaptive estimators over the collection of anisotropic Nikol'skii classes $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ under $\mathbb{L}_p$-loss.

The last advances in this task were made in Goldenshluger and Lepski (2014). However, as was conjectured in this paper, their local approach cannot lead to the construction of optimally-adaptive estimators. On the other hand it is not clear how to adapt the approach developed in the present paper to the density estimation on $\mathbb{R}^d$. Indeed, the key element of our procedure are the upper functions for $\mathbb{L}_p$-norm of random fields found in Lepski (2015). These results are heavily based on the fact that the corresponding norm is defined on a bounded interval of $\mathbb{R}^d$.

The same problem can be formulated for the more complicated *deconvolution model*. Recent advances in the estimation under $\mathbb{L}_2$-loss in this model can be found in Comte and Lacour (2013).

4.4. *Regression model.* Let $\xi_i, i \in \mathbb{N}^*$, be *i.i.d. symmetric* random variables with common probability density ϱ, and let $X_i, i \in \mathbb{N}^*$, be *i.i.d.* d-dimensional random vectors with common probability density g. Suppose that we observe the pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ satisfying

$$Y_i = f(X_i) + \xi_i, \qquad i = 1, \dots, n.$$

The goal is to estimate function f under $\mathbb{L}_p$-loss on $(-b, b)^d$, where $b > 0$ is a given number.

We will suppose that the sequences $\xi_i, i \in \mathbb{N}^*$ and $X_i, i \in \mathbb{N}^*$ are mutually independent and that the design density g (known or unknown) is separated away from zero on $(-b, b)^d$.

Regular noise. Suppose that there exists $a > 0$ and $A > 0$ such that for any $u, v \in [-a, a]$

$$\int_{\mathbb{R}} \frac{\varrho(y+u)\varrho(y+v)}{\varrho(y)} \, \mathrm{d}y \le 1 + A|uv|. \tag{4.1}$$

Assume also that $\mathbb{E}|\xi_1|^{\alpha} < \infty$ for some $\alpha \ge 2$.

PROBLEM V. Prove or disprove the existence of optimally-adaptive estimators over the collection of anisotropic Nikol'skii classes $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ under $\mathbb{L}_p$-loss.

The interesting question arising in this context is what is the minimal value of α under which the formulated problem can be solved. In particular, is it or is it not related to the norm index p?

Cauchy noise. Let $\varrho(x) = \{\pi(1+x^2)\}^{-1}$. In this case the noise is of course regular; that is, (4.1) holds, but the moment assumption fails. To our knowledge, there are no minimax or minimax adaptive results in the multivariate regression model with noise "without moments" when anisotropic functional classes are considered.

PROBLEM VI. Propose the construction of optimally-adaptive estimators over the scale of anisotropic Hölder classes $\mathbb{N}_{\vec{\infty},d}(\vec{\beta}, \vec{L})$ under $\mathbb{L}_p$-loss.

The same problem can be of course formulated over the scale of anisotropic Nikol'skii classes, but it seems that nowadays neither probabilistic nor the tools from functional analysis are sufficiently developed in order to proceed to this task.

Irregular noise. Consider the following two particular examples:

$$\varrho(x) = 2^{-1} 1_{[-1,1]}(x), \qquad \varrho_\gamma(x) = C_\gamma e^{-|x|^\gamma}, \qquad \gamma \in (0, 1/2).$$

In both cases, condition (4.1) is not fulfilled. In the parametric case $f(\cdot) \equiv \mathbf{f} \in \mathbb{R}$, the minimax rate of convergence is faster than $n^{-1/2}$; see Ibragimov and Has'minskiĭ (1981).

PROBLEM VII. Find minimax rate of convergence on anisotropic Hölder class $\mathbb{N}_{\vec{\infty},d}(\vec{\beta}, \vec{L})$ under $\mathbb{L}_p$-loss. Propose an aggregation scheme for these estimators that leads to the construction of optimally-adaptive estimators.

One of the possible approaches to solving Problems VI and VII could be an $\mathbb{L}_p$-aggregation of locally-Bayesian or M-estimators. Some recent advances in this direction can be found in Chichignoud (2012), Chichignoud and Lederer (2014).

Unknown distribution of the noise. Suppose now that the density ϱ is unknown or even does not exist. The goal is to consider simultaneously the noises with and without moments, regular or irregular etc. Even if the regression function belong to *known* functional class, the different noises may lead to different minimax rates of convergence.

PROBLEM VIII. Build an estimator which would simultaneously adapt to a given scale of functional classes and to the noise distribution.

We do not precise here the collection of classes since the formulated problem seems extremely complicated. Any solution, even in dimension 1, can be considered as a great progress. In this context let us mention some very promising results recently obtained in Baraud, Birgé and Sart (2014).

We finish this section with following remarks. The regression model is very rich, and many other problems can be formulated from its framework. For instance, the discussed problems can be mixed with imposing structural assumptions on the model. On the other hand aforementioned problems are not directly related to the concrete statistical model. In particular, almost all of them can be postulated in the *inverse problem* estimation context or in nonparametric auto-regression.

5. Proof of Theorem 1 and Corollary 1. We start this section by presenting the constants appearing in the assertion of Theorem 1.

5.1. *Important quantities.* Put

$$
\begin{aligned}
C_1 &= 2\big(q \vee \big[p1\{p<\infty\} + 1\{p=\infty\}\big]\big) \\
&\qquad + 2\sqrt{2d}\big[\sqrt{\pi} + \|K\|_2\big(\sqrt{\big|\ln\big(4bA\|K\|_2\big)\big|} + 1\big)\big]; \\
C_3 &= C_3(\tilde{q}, p)1\{p<\infty\} + C_3(q,1)1\{p=\infty\}, \qquad \tilde{q} = (q/p) \vee 1; \\
C_4 &= \bigg(\boldsymbol{\gamma}_{q+1}\sqrt{(\pi/2)}\big[1 \vee (2b)^{qd}\big] \sum_{r\in\mathbb{N}_p^*} e^{-e^r}\big[(r\sqrt{e})^d \|\mathcal{K}\|_{2r/(r+2)}^d\big]^{q/2}\bigg)^{1/q},
\end{aligned}
$$

where $\boldsymbol{\gamma}_{q+1}$ is the $(q+1)$th absolute moment of the standard normal distribution and

$$
C_3(u,v) = (4b)^{d/v}\bigg[2u\int_0^\infty z^{u-1}\exp\bigg(-\frac{z^{2/v}}{8\|K\|_2^2}\bigg)\,\mathrm{d}z\bigg]^{1/(uv)}, \qquad u, v \geq 1.
$$

5.2. *Auxiliary results.* As we have mentioned, the main ingredient of the proof of Theorem 1 is the fact that $\{\Psi_\varepsilon(\vec{h}), \vec{h} \in \mathbb{H}\}$ is the upper function for the collection $\{\|\xi_{\vec{h}}\|_p, \vec{h} \in \mathbb{H}\}$. The corresponding result is formulated below for citation convenience, as Proposition 1 is proved in Theorem 1 and in Corollary 2 of Theorem 2, Lepski (2015).

Set for any $p \in [1,\infty)$, $\tau \in (0,1)$ and $\mathcal{L} > 0$,

$$
\psi_\varepsilon(\vec{h}) = \widetilde{\Psi}_{\varepsilon,p}(\vec{h}) \wedge \Big(\inf_{r\in\mathbb{N}_p^*(\vec{h},\mathcal{A}_\varepsilon)} C_2(r,\tau,\mathcal{L})\big\|V_{\vec{h}}^{-1/2}\big\|_{rp/(r-p)}\Big), \qquad \vec{h} \in \mathbb{B}(\mathcal{A}_\varepsilon).
$$

PROPOSITION 1. *Let $\mathcal{L} > 0$ be fixed, and let $\mathfrak{h}_\varepsilon$ and $\mathcal{A}_\varepsilon$ be defined in* (2.2). *Suppose also that K satisfies Assumption* 1.

Then for any $q \geq 1$ and $\tau \in (0,1)$ one can find $\varepsilon(\tau, q)$ such that:

(1) *for any $p \in [1,\infty)$, $\varepsilon \leq \varepsilon(\tau,q)$ and any countable $\mathrm{H} \subset \mathfrak{S}_d(\mathfrak{h}_\varepsilon) \cap \mathbb{H}_d(\tau, \mathcal{L}, \mathcal{A}_\varepsilon)$, one has*

$$\mathbb{E}\Big\{\sup_{\vec{h}\in \mathrm{H}}\big[\|\xi_{\vec{h}}\|_p - \psi_\varepsilon(\vec{h})\big]_+\Big\}^q \leq \{(C_3 + C_4)\varepsilon\}^q;$$

(2) *for any $p \in [1,\infty]$, $\varepsilon \leq \varepsilon(\tau,q)$ and any countable $\mathrm{H} \subset \mathfrak{S}_d$,*

$$\mathbb{E}\Big\{\sup_{\vec{h}\in \mathrm{H}}\big[\|\xi_{\vec{h}}\|_p - \widetilde{\Psi}_{\varepsilon,p}(\vec{h})\big]_+\Big\}^q \leq \{C_3\varepsilon\}^q.$$

We will need also the following technical result.

LEMMA 1. *For any $d \geq 1$, $\varkappa \in (0, 1/d)$, $\mathfrak{L} > 0$ and $\mathcal{A} \geq e^d$*:

(i) $\quad \mathbb{H}_d(\varkappa, \mathfrak{L}, \mathcal{A}) \subseteq \mathbb{H}_d(d\varkappa, \mathfrak{L}^d, \mathcal{A})$;

(ii) $\quad \vec{h} \vee \vec{\eta} \in \mathbb{H}_d(d\varkappa, (2\mathfrak{L})^d, \mathcal{A}) \qquad \forall \vec{h}, \vec{\eta} \in \mathbb{H}_d(\varkappa, \mathfrak{L}, \mathcal{A})$.

The first statement of the lemma is obvious, and the second will be proved in the Appendix.

5.3. *Proof of Theorem* 1. Let $\vec{h} \in \mathbb{H}$ be fixed. We have in view of the triangle inequality

$$\|\hat{f}_{\vec{\mathbf{h}}} - f\|_p \leq \|\hat{f}_{\vec{\mathbf{h}}\vee\vec{h}} - \hat{f}_{\vec{\mathbf{h}}}\|_p + \|\hat{f}_{\vec{\mathbf{h}}\vee\vec{h}} - \hat{f}_{\vec{h}}\|_p + \|\hat{f}_{\vec{h}} - f\|_p. \tag{5.1}$$

First, note that $\hat{f}_{\vec{\mathbf{h}}\vee\vec{h}} \equiv \hat{f}_{\vec{h}\vee\vec{\mathbf{h}}}$ and, therefore,

$$\begin{aligned}\|\hat{f}_{\vec{\mathbf{h}}\vee\vec{h}} - \hat{f}_{\vec{\mathbf{h}}}\|_p = \|\hat{f}_{\vec{h}\vee\vec{\mathbf{h}}} - \hat{f}_{\vec{\mathbf{h}}}\|_p &\leq \widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) + \varepsilon\Psi_{\varepsilon,p}(\vec{h}\vee\vec{\mathbf{h}}) + \varepsilon\Psi_{\varepsilon,p}(\vec{\mathbf{h}})\\ &\leq \widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) + 2\varepsilon\Psi_{\varepsilon,p}(\vec{\mathbf{h}}).\end{aligned} \tag{5.2}$$

Here we used first $\vec{\mathbf{h}} \in \mathbb{H}$ and then that $V_{\vec{h}\vee\vec{\eta}} \geq V_{\vec{h}} \vee V_{\vec{\eta}}$ implies $\Psi_{\varepsilon,p}(\vec{h}\vee\vec{\eta}) \leq \Psi_{\varepsilon,p}(\vec{h}) \wedge \Psi_{\varepsilon,p}(\vec{\eta})$ for any $\vec{h}$ and $\vec{\eta}$. Similarly we have

$$\begin{aligned}\|\hat{f}_{\vec{\mathbf{h}}\vee\vec{h}} - \hat{f}_{\vec{h}}\|_p &\leq \widehat{\mathcal{R}}_{\mathbb{H}}(\vec{\mathbf{h}}) + \varepsilon\Psi_{\varepsilon,p}(\vec{h}\vee\vec{\mathbf{h}}) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\\ &\leq \widehat{\mathcal{R}}_{\mathbb{H}}(\vec{\mathbf{h}}) + 2\varepsilon\Psi_{\varepsilon,p}(\vec{h}).\end{aligned} \tag{5.3}$$

The definition of $\vec{\mathbf{h}}$ implies

$$\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) + 2\varepsilon\Psi_{\varepsilon,p}(\vec{h}) + \widehat{\mathcal{R}}_{\mathbb{H}}(\vec{\mathbf{h}}) + 2\varepsilon\Psi_{\varepsilon,p}(\vec{\mathbf{h}}) \leq 4\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) + 4\varepsilon\Psi_{\varepsilon,p}(\vec{h}) + 2\varepsilon,$$

and we get from (5.1), (5.2) and (5.3)

$$\|\hat{f}_{\vec{\mathbf{h}}} - f\|_p \leq 4\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) + 4\varepsilon\Psi_{\varepsilon,p}(\vec{h}) + \|\hat{f}_{\vec{h}} - f\|_p + 2\varepsilon. \tag{5.4}$$

We obviously have for any $\vec{h}, \vec{\eta} \in \mathbb{H}$,

$$\|\hat{f}_{\vec{h}\vee\vec{\eta}} - \hat{f}_{\vec{\eta}}\|_p \le \|B_{\vec{h},\vec{\eta}}\|_p + \varepsilon\|\xi_{\vec{h},\vec{\eta}}\|_p + \varepsilon\|\xi_{\vec{\eta}}\|_p.$$

Denote $\mathbb{H}^* = \{\vec{v} : \vec{v} = \vec{h} \vee \vec{\eta}, \vec{h}, \vec{\eta} \in \mathbb{H}\}$, and remark that $\mathbb{H} \subseteq \mathbb{H}^* \subseteq \mathbb{H}(d\varkappa, (2\mathfrak{L})^d, \mathcal{A}_\varepsilon)$. The latter inclusion follows from assertions of Lemma 1. Moreover since $\mathbb{H}$ is countable, $\mathbb{H}^*$ is countable as well. Putting $\zeta = \sup_{\vec{v}\in\mathbb{H}^*}[\|\xi_{\vec{v}}\|_p - \Psi_{\varepsilon,p}(\vec{v})]_+$, we obtain

$$\widehat{\mathcal{R}}_{\mathbb{H}}(\vec{h}) \le \sup_{\vec{\eta}\in\mathbb{H}} \|B_{\vec{h},\vec{\eta}}(\cdot, f)\|_p + 2\varepsilon\zeta$$

and therefore, in view of (5.4),

$$\|\hat{f}_{\vec{\mathbf{h}}} - f\|_p \le 4 \sup_{\vec{\eta}\in\mathbb{H}} \|B_{\vec{h},\vec{\eta}}(\cdot, f)\|_p + 4\varepsilon\Psi_{\varepsilon,p}(\vec{h}) + 8\varepsilon\zeta + \|\hat{f}_{\vec{h}} - f\|_p + 2\varepsilon.$$

Taking into account that $\|\hat{f}_{\vec{h}} - f\|_p \le \|B_{\vec{h}}\|_p + \varepsilon\|\xi_{\vec{h}}\|_p$, we obtain

$$\|\hat{f}_{\vec{\mathbf{h}}} - f\|_p \le 5\mathcal{B}^{(p)}_{\vec{h}}(f) + 5\varepsilon\Psi_{\varepsilon,p}(\vec{h}) + 9\varepsilon\zeta + 2\varepsilon.$$

It remains to note that if $p \in [1, \infty)$ in view of the definition of $\Psi_{\varepsilon,p}(\cdot)$,

$$\zeta = \Big(\sup_{\vec{v}\in\mathbb{H}^*\cap\mathfrak{S}_d(\mathfrak{h}_\varepsilon)} [\|\xi_{\vec{v}}\|_p - \overline{\Psi}_{\varepsilon,p}(\vec{v})]_+\Big) \vee \Big(\sup_{\vec{v}\in\mathbb{H}^*\setminus\mathfrak{S}_d(\mathfrak{h}_\varepsilon)} [\|\xi_{\vec{v}}\|_p - \widetilde{\Psi}_{\varepsilon,p}(\vec{v})]_+\Big).$$

Applying the first and the second assertions of Proposition 1 with $\tau = d\varkappa$, $\mathcal{L} = (2\mathfrak{L})^d$, $\mathrm{H} = \mathbb{H}^* \cap \mathfrak{S}_d(\mathfrak{h}_\varepsilon)$ and $\mathrm{H} = \mathbb{H}^* \setminus \mathfrak{S}_d(\mathfrak{h}_\varepsilon)$, respectively, we obtain

$$\mathcal{R}^{(p)}_\varepsilon[\tilde{f}_{\vec{\mathbf{h}}}; f] \le 5\mathcal{B}^{(p)}_{\vec{h}}(f) + 5\varepsilon\Psi_{\varepsilon,p}(\vec{h}) + 18(C_3 + C_4 + 2)\varepsilon.$$

It remains to note that the left-hand side of the obtained inequality is independent of $\vec{h}$, and we come to the assertion of the theorem with $\Upsilon = 18(C_3 + C_4 + 2)$, where, recall, C_3 and C_4 are given in Section 5.1.

If $p = \infty$, the second assertion of Proposition 1 with $\mathrm{H} = \mathbb{H}^*$ is directly applied to the random variable ζ, and the statement of the theorem follows.

5.4. *Proof of Corollary* 1. The proof of the corollary consists mostly of bounding from above the quantity $\mathcal{B}^{(p)}_{\vec{h}}(f)$. This, in its turn, is based on the technical result presented in Lemma 2 below, which will be used in the proof of Proposition 2, Section 7.3.1, as well.

5.4.1. *Auxiliary lemma*. The following notation will be exploited in the sequel:

For any $J \subseteq \{1, \dots, d\}$ and $y \in \mathbb{R}^d$, set $y_J = \{y_j, j \in J\} \in \mathbb{R}^{|J|}$, and we will write $y = (y_J, y_{\bar{J}})$, where as usual $\bar{J} = \{1, \dots, d\} \setminus J$.

For any $j = 1, \dots, d$, introduce $\mathbf{E}_j = (\mathbf{0}, \dots, \mathbf{e}_j, \dots, \mathbf{0})$, and set $\mathbf{E}[J] = \sum_{j\in J} \mathbf{E}_j$. Later on $\mathbf{E}_0 = \mathbf{E}[\varnothing]$ denotes the matrix with zero entries.

To any $J \subseteq \{1,\dots,d\}$ and any $\lambda:\mathbb{R}^d \to \mathbb{R}_+$ such that $\lambda \in \mathbb{L}_p(\mathbb{R}^d)$, associate the function

$$\lambda(y_J, z_{\bar{J}}) = \lambda\bigl(z + \mathbf{E}[J](y-z)\bigr), \qquad y, z \in \mathbb{R}^d,$$

with the obvious agreement $\lambda_J \equiv \lambda$ if $J = \{1,\dots,d\}$, which is always the case if $d=1$.

At last for any $\vec{\mathrm{h}} = (\mathrm{h}_1,\dots,\mathrm{h}_d) \in \mathfrak{S}_d^{\mathrm{const}}$ and $J \subseteq \{1,\dots,d\}$, set $K_{\vec{\mathrm{h}},J}(u_J) = \prod_{j\in J} \mathrm{h}_j^{-1}\mathcal{K}(u_j/\mathrm{h}_j)$, and define for any $y \in \mathbb{R}^d$,

$$[K_{\vec{\mathrm{h}}} \star \lambda]_J(y) = \int_{\mathbb{R}^{|\bar{J}|}} K_{\vec{\mathrm{h}},\bar{J}}(u_{\bar{J}} - y_{\bar{J}})\lambda(y_J, u_{\bar{J}})\nu_{|\bar{J}|}(\mathrm{d}u_{\bar{J}}).$$

The following result is a trivial consequence of the Young inequality and Fubini's theorem. For any $J \subseteq \{1,\dots,d\}$ and $p \in [1,\infty]$,

$$\bigl\|[K_{\vec{\mathrm{h}}} \star \lambda]_J\bigr\|_p \le \|\mathcal{K}\|_{1,\mathbb{R}}^{d-|J|}\|\lambda\|_{p,\mathcal{A}_J} \qquad \forall \vec{\mathrm{h}} \in \mathfrak{S}_d^{\mathrm{const}}, \tag{5.5}$$

where we have denoted $\mathcal{A}_J = (-b,b)^{|J|} \times \mathbb{R}^{|\bar{J}|}$.

LEMMA 2. *For any $\vec{h}, \vec{\eta} \in \mathfrak{S}_d^{\mathrm{const}}$, one can find $k = 1,\dots,d$, and the collection of indexes $\{j_1 < j_2 < \cdots < j_k\} \in \{1,\dots,d\}$ such that for any $x \in \mathbb{R}^d$ and any $f:\mathbb{R}^d \to \mathbb{R}$,*

$$B_{\vec{h},\vec{\eta}}(x,f) \le \sum_{l=1}^{k}\bigl(\bigl[|K_{\vec{h}\vee\vec{\eta}}| \star b_{\vec{h},j_l}\bigr]_{\mathcal{J}_l}(x) + \bigl[|K_{\vec{\eta}}| \star b_{\vec{h},j_l}\bigr]_{\mathcal{J}_l}(x)\bigr);$$

$$B_{\vec{h}}(x,f) \le \sum_{l=1}^{k}\bigl[|K_{\vec{h}}| \star b_{\vec{h},j_l}\bigr]_{\mathcal{J}_l}(x), \qquad \mathcal{J}_l = \{j_1,\dots,j_l\}.$$

The proof of the lemma is postponed to the Appendix.

5.4.2. *Proof of the corollary.* We obtain in view of the first assertion of Lemma 2 and (5.5),

$$\bigl\|B_{\vec{h},\vec{\eta}}(\cdot,f)\bigr\|_p \le 2\sum_{l=1}^{k}\|\mathcal{K}\|_{1,\mathbb{R}}^{d-l}\|b_{\vec{h},j_l}\|_{p,\mathcal{A}_{J_l}} = 2\sum_{l=1}^{k}\|\mathcal{K}\|_{1,\mathbb{R}}^{d-l}\|b_{\vec{h},j_l}\|_p.$$

The latter equality follows from the fact that f is compactly supported on $(-b,b)^d$, which implies that $b_{\vec{h},j_l}(x_{\mathcal{J}_l},\cdot)$ is compactly supported on $(-b,b)^{d-l}$. Taking into account that $\|\mathcal{K}\|_{1,\mathbb{R}} \ge 1$, we get

$$\bigl\|B_{\vec{h},\vec{\eta}}(\cdot,f)\bigr\|_p \le 2\|\mathcal{K}\|_{1,\mathbb{R}}^{d}\sum_{l=1}^{k}\|b_{\vec{h},j_l}\|_p \le 2\|\mathcal{K}\|_{1,\mathbb{R}}^{d}\sum_{j=1}^{d}\|b_{\vec{h},j}\|_p \qquad \forall \vec{h}, \vec{\eta} \in \mathfrak{S}_d^{\mathrm{const}}.$$

Since the right-hand side of the latter inequality is independent of $\vec{\eta}$, we obtain

$$\sup_{\vec{\eta}\in\mathfrak{S}_d^{\mathrm{const}}} \|B_{\vec{h},\vec{\eta}}(\cdot, f)\|_p \le 2\|\mathcal{K}\|_{1,\mathbb{R}}^d \sum_{j=1}^d \|b_{\vec{h},j}\|_p.$$

Repeating previous computations and using the second assertion of Lemma 2, we have

$$\|B_{\vec{h}}(\cdot, f)\|_p \le \|\mathcal{K}\|_{1,\mathbb{R}}^d \sum_{j=1}^d \|b_{\vec{h},j}\|_p \tag{5.6}$$

for any $\vec{h} \in \mathfrak{S}_d^{\mathrm{const}}$ and any $p \ge 1$. We obtain finally

$$\mathcal{B}_{\vec{h}}^{(p)}(f) \le 3\|\mathcal{K}\|_{1,\mathbb{R}}^d \sum_{j=1}^d \|b_{\vec{h},j}\|_p. \tag{5.7}$$

3^0. We obviously have $r_{\mathcal{A}}(\vec{h}) = \lfloor p \rfloor + 1$ for any $\vec{h} \in \mathbb{H}_\varepsilon$, and therefore, for any $p \in [1, \infty)$,

$$\Psi_{\varepsilon,p}(\vec{h}) \le (2b)^{d/p} V_{\vec{h}}^{-1/2} \inf_{r\in\mathbb{N}_p^*} C_2(r) = \Psi_{\varepsilon,p}^{(\mathrm{const})}(\vec{h}).$$

It is also obvious that

$$\Psi_{\varepsilon,\infty}(\vec{h}) = \Psi_{\varepsilon,\infty}^{(\mathrm{const})}(\vec{h}) \qquad \forall \vec{h} \in \mathbb{H}_\varepsilon.$$

As we previously mentioned $\mathfrak{S}_d^{\mathrm{const}} \subset \mathbb{H}_d(\varkappa, \mathfrak{L})$ for any $\varkappa \in (0, 1)$ and $\mathfrak{L} = (2b)^{\varkappa}$. Thus, choosing, for example, $\varkappa = (2d)^{-1}$, we get that $\mathbb{H}_\varepsilon^{\mathrm{const}} \subset \mathbb{H}_d((2d)^{-1}, (2b)^{1/(2d)}, \mathcal{A}_\varepsilon)$, and, moreover, $\mathbb{H}_\varepsilon^{\mathrm{const}}$ is obviously finite set.

The assertion of the corollary follows now from (5.7) and Theorem 1.

6. Proof of Theorem 2. The proof is organized as follows. First, we formulate two auxiliary statements, Lemmas 3 and 4. Second, we present a general construction of a finite set of functions employed in the proof of lower bounds. Then we specialize the constructed set of functions in different regimes and derive the announced lower bounds.

6.1. *Proof of Theorem* 2. *Auxiliary lemmas*. The first statement given in Lemma 3 is a simple consequence of Theorem 2.4 from Tsybakov (2009). Let $\mathbb{F}$ be a given set of real functions defined on $(-b, b)^d$.

LEMMA 3. *Assume that for any sufficiently small* $\varepsilon > 0$, *one can find a positive real number* ρ_ε *and a finite subset of functions* $\{f^{(0)}, f^{(j)}, j \in \mathcal{J}_\varepsilon\} \subset \mathbb{F}$ *such that*

$$\|f^{(i)} - f^{(j)}\|_p \ge 2\rho_\varepsilon \qquad \forall i, j \in \mathcal{J}_\varepsilon \cup \{0\} : i \ne j; \tag{6.1}$$

$$\limsup_{\varepsilon\to 0} \frac{1}{|\mathcal{J}_\varepsilon|^2} \sum_{j\in\mathcal{J}_\varepsilon} \mathbb{E}_{f^{(0)}} \left\{ \frac{\mathrm{d}\mathbb{P}_{f^{(j)}}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}(X^{(\varepsilon)}) \right\}^2 =: C < \infty. \tag{6.2}$$

Then for any $q \geq 1$,

$$\liminf_{\varepsilon\to 0} \inf_{\tilde{f}} \sup_{f\in\mathbb{F}} \rho_\varepsilon^{-1}\big(\mathbb{E}_f \|\tilde{f} - f\|_p^q\big)^{1/q} \geq (\sqrt{C} + \sqrt{C+1})^{-2/q},$$

where infimum on the left-hand side is taken over all possible estimators.

We will apply Lemma 3 with $\mathbb{F} = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$.

Next, we will need the result to be a generalization of the Varshamov–Gilbert lemma. It can be found in Rigollet and Tsybakov (2011), Lemma A3. In the version established in Lemma 4 below, we only provide a particular choice of the constants appearing in the latter result.

Let ϱ_n be the Hamming distance on $\{0, 1\}^n$, $n \in \mathbb{N}^*$, that is,

$$\varrho_n(a, b) = \sum_{j=1}^n \mathbf{1}\{a_j \neq b_j\} = \sum_{j=1}^n |a_j - b_j|, \qquad a, b \in \{0, 1\}^n.$$

LEMMA 4. *For any* $m \geq 4$ *there exist a subset* $\mathcal{P}_{m,n}$ *of* $\{0, 1\}^n$ *such that*

$$|\mathcal{P}_{m,n}| \geq 2^{-m}(n/m - 1)^{m/2}, \qquad \sum_{k=1}^n a_k = m, \qquad \varrho_m(a, a') \geq m/2 \qquad \forall a, a' \in \mathcal{P}_{m,n}.$$

6.2. *Proof of Theorem* 2. *General construction of a finite set of functions.* This part of the proof is mostly based on the constructions and computations made in Goldenshluger and Lepski (2014), proof of Theorem 3. For any $t \in \mathbb{R}$, set

$$g(t) = e^{-1/(1-t^2)}\mathbf{1}_{[-1,1]}(t).$$

For any $l = 1, \dots, d$ let $b/2 > \sigma_l = \sigma_l(\varepsilon) \to 0$, $\varepsilon \to 0$ be the sequences to be specified later. Let $M_l = \sigma_l^{-1}$, and without loss of generality assume that M_l, $l = 1, \dots, d$ are integer numbers.

Define also

$$x_{j,l} = -b + 2j\sigma_l, \qquad j = 1, \dots, M_l, l = 1, \dots, d,$$

and let $\mathcal{M} = \{1, \dots, M_1\} \times \cdots \times \{1, \dots, M_d\}$. For any $\mathfrak{m} = (\mathfrak{m}_1, \dots, \mathfrak{m}_d) \in \mathcal{M}$, define

$$\pi(\mathfrak{m}) = \sum_{j=1}^{d-1}(\mathfrak{m}_j - 1)\left(\prod_{l=j+1}^d M_l\right) + \mathfrak{m}_d,$$

$$G_{\mathfrak{m}}(x) = \prod_{l=1}^d g\left(\frac{x_l - x_{\mathfrak{m}_l,l}}{\sigma_l}\right), \qquad x \in \mathbb{R}^d.$$

Let W be a subset of $\{0,1\}^{|\mathcal{M}|}$. Define a family of functions $\{f_w, w\in W\}$ by

$$f_w(x)=A\sum_{\mathfrak{m}\in\mathcal{M}} w_{\pi(\mathfrak{m})}G_{\mathfrak{m}}(x),\qquad x\in\mathbb{R}^d,$$

where w_j, $j=1,\dots,|\mathcal{M}|$ are the coordinates of w, and A is a parameter to be specified.

Suppose that the set W is chosen so that

$$\varrho_{|\mathcal{M}|}(w,w')\ge B\qquad \forall w,w'\in W, \tag{6.3}$$

where we remind the reader that $\varrho_{|\mathcal{M}|}$ is the Hamming distance on $\{0,1\}^{|\mathcal{M}|}$. Here $B=B(\varepsilon)\ge 1$ is a parameter to be specified. Let also $S_W:=\sup_{w\in W}|\{j: w_j\neq 0\}|$. Note finally that f_w, $w\in W$, are compactly supported on $(-b,b)^d$.

Repeating the computations made in Goldenshluger and Lepski (2014), proof of Theorem 3, we assert first that if

$$A\sigma_l^{-\beta_l}\Bigg(S_W\prod_{j=1}^d\sigma_j\Bigg)^{1/r_l}\le C_1^{-1}L_l\qquad \forall l=1,\dots,d, \tag{6.4}$$

then $f_w\in\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})$ for any $w\in W$. Here C_1 as well as C_2 and C_3 defined below are the numerical constants completely determined by the function g.

Next the condition (6.1) of Lemma 3 is fulfilled with

$$\rho_\varepsilon=C_2A\Bigg(B\prod_{j=1}^d\sigma_j\Bigg)^{1/p}, \tag{6.5}$$

which remains true if $p=\infty$ as well. At last, we have $\|f_w\|_2^2\le C_3A^2S_W\prod_{j=1}^d\sigma_j$.

Set $f^{(0)}\equiv 0$, and let us verify condition (6.2) of Lemma 3. First observe that in view of Girsanov's formulas,

$$\frac{\mathrm{d}\mathbb{P}_{f_w}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\big(X^{(\varepsilon)}\big)=\exp\bigg\{\varepsilon^{-1}\int f_w b(\mathrm{d}t)-\big(2\varepsilon^2\big)^{-1}\|f_w\|_2^2\bigg\}.$$

This yields that for any $w\in W$,

$$\mathbb{E}_{f^{(0)}}\bigg\{\frac{\mathrm{d}\mathbb{P}_{f_w}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\big(X^{(\varepsilon)}\big)\bigg\}^2=\exp\big\{\varepsilon^{-2}\|f_w\|_2^2\big\}\le\exp\Bigg\{\varepsilon^{-2}C_3A^2S_W\prod_{j=1}^d\sigma_j\Bigg\}.$$

The right-hand side of the latter inequality does not depend on w; hence we have

$$\frac{1}{|W|^2}\sum_{w\in W}\mathbb{E}_{f^{(0)}}\bigg\{\frac{\mathrm{d}\mathbb{P}_{f_w}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\big(X^{(\varepsilon)}\big)\bigg\}^2\le\exp\Bigg\{C_3\varepsilon^{-2}A^2S_W\Bigg(\prod_{j=1}^d\sigma_j\Bigg)-\ln\big(|W|\big)\Bigg\}.$$

Therefore, condition (6.2) of Lemma 3 is fulfilled with $C=1$ if

$$C_3\varepsilon^{-2}A^2S_W\prod_{j=1}^d\sigma_j\le\ln\big(|W|\big). \tag{6.6}$$

In order to apply Lemma 3 it remains to specify the parameters A, σ_l, $l = 1,\dots,d$ and the set W so that relationships (6.3), (6.4) and (6.6) are simultaneously fulfilled. According to Lemma 3, under these conditions the lower bound is given by ρ_ε in (6.5).

6.3. *Proof of Theorem 2. Choice of the parameters.* We begin with the construction of the set W. Let $m \geq 4$ be an integer number whose choice will be made later, and, without loss of generality, assume that $|\mathcal{M}|/m \geq 9$ is integer. Let $\mathcal{P}_{m,|\mathcal{M}|}$ be a subset of $\{0,1\}^{|\mathcal{M}|}$ defined in Lemma 4, where we put $n = |\mathcal{M}|$.

Set $W = \mathcal{P}_{m,|\mathcal{M}|} \cup \mathbf{0}$, where $\mathbf{0}$ is the zero sequence of the size $|\mathcal{M}|$. With such a set W,

$$S_W \leq m, \qquad \ln(|W|) \geq (m/2)\big[\ln_2(|\mathcal{M}|/m - 1) - 2\big],$$

and, therefore, condition (6.6) holds true if

$$A^2\varepsilon^{-2}\prod_{j=1}^{d}\sigma_j \leq (2C_3)^{-1}\big[\ln_2(|\mathcal{M}|/m - 1) - 2\big]. \tag{6.7}$$

We also note that condition (6.4) is fulfilled if we require

$$A\sigma_l^{-\beta_l}\left(m\prod_{j=1}^{d}\sigma_j\right)^{1/r_l} \leq C_1^{-1}L_l \qquad \forall l = 1,\dots,d. \tag{6.8}$$

In addition, (6.3) holds with $B = m/2$, and therefore,

$$\rho_\varepsilon = 2^{-1/p}C_2A\left(m\prod_{j=1}^{d}\sigma_j\right)^{1/p}. \tag{6.9}$$

6.4. *Proof of Theorem 2. Derivation of lower bounds in different zones.* Let $\mathbf{c}_i$, $\mathbf{i} = 1,\dots,6$, be constants; the choice will be made later.

Case: $\varkappa(p) \leq 0$, $\tau(p^*) \leq 0$. Set

$$\varpi_\varepsilon = \begin{cases} \big(L_\beta\varepsilon^2|\ln(\varepsilon)|\big)^{\omega/(\varkappa(p^*))}, & \varkappa(p^*) < 0; \\ L_\beta e^{-\varepsilon^{-2}}, & \varkappa(p^*) = 0, \end{cases}$$

and note that $\varpi_\varepsilon \to \infty$, $\varepsilon \to 0$. In view of the latter remark we will assume that ε is small enough, provided $\varpi_\varepsilon > 1$. We start our considerations with the following remark. The case $\varkappa(p^*) = 0$ is possible only if $p^* = p$ since $\varkappa(\cdot)$ is strictly decreasing. Moreover, in view of relation (3.3), $\varkappa(p^*) = 0$ is possible only if $p \leq 2$ since $\tau(p^*) \leq 0$. Choose

$$A = \mathbf{c}_1\varpi_\varepsilon, \qquad m = \mathbf{c}_2L_\beta\varpi_\varepsilon^{-p^*\tau(p^*)}, \qquad \sigma_l = \mathbf{c}_3L_l^{-1/\beta_l}\varpi_\varepsilon^{(r_l-p^*)/(\beta_lr_l)}.$$

With this choice, we have $\frac{|\mathcal{M}|}{m} = m^{-1}\prod_{j=1}^{d}\sigma_j^{-1} = \mathbf{c}_2^{-1}\mathbf{c}_3^{-d}\varpi_\varepsilon^{p^*} \to \infty, \varepsilon \to 0$. Hence, for any ε small enough, one has

$$[\ln_2(|\mathcal{M}|/m - 1) - 2] \geq Q_1 \begin{cases} |\ln(\varepsilon)|, & \varkappa(p^*) < 0; \\ \varepsilon^{-2}, & \varkappa(p^*) = 0, \end{cases}$$

where Q_1 is independent of ε and $\vec{L}$. This yields that (6.7) and (6.8) will be fulfilled if

$$\mathbf{c}_1^2\mathbf{c}_3^d \leq (2C_3)^{-1}Q_1, \qquad \mathbf{c}_1\mathbf{c}_2^{1/r_l}\mathbf{c}_3^{d/r_l-\beta_l} \leq C_1^{-1}. \tag{6.10}$$

Some remarks are in order. First, since $r_l \leq p^*$ for any $l = 1, \dots, d$ and $\varkappa(p^*) < 0$, we have

$$\sigma_l \leq \mathbf{c}_3 L_l^{-1/\beta_l} \leq \mathbf{c}_3\Big[\min_{l=1,\dots,d} L_0^{-1/\beta_l}\Big].$$

Here we also used $\varpi_\varepsilon > 1$. Thus choosing $\mathbf{c}_3$ small enough we can guarantee that $\sigma_l \leq b/2$ for any $l = 1, \dots, d$, which was the unique restriction imposed on the choice of the latter sequence.

Next $\tau(p^*) \leq 0$, $\varpi_\varepsilon > 1$ and $p^*\tau(p^*) = 2 - p^*$, when $\varkappa(p^*) = 0$, imply that $m \geq \mathbf{c}_2 L_0^{1/\beta}$, and therefore, choosing $\mathbf{c}_2$ large enough we guarantee that $m \geq 4$. At last, choosing $\mathbf{c}_1$ small enough we can assert that (6.10) is satisfied.

Thus it remains to compute ρ_ε. We get from (6.9)

$$\begin{aligned} \rho_\varepsilon &= C_2 2^{-1/p}\mathbf{c}_1(\mathbf{c}_2\mathbf{c}_3^d)^{1/p}\varpi_\varepsilon^{1-p^*/p} \\ &=: \mathbf{1}_{(p,\infty]}(p^*)Q_2(L_\beta\varepsilon^2|\ln(\varepsilon)|)^{\omega(p^*-p)/(p(p^*-\omega(2+1/\beta)))}. \end{aligned} \tag{6.11}$$

We remark that there are no uniformly consistent estimators if $p^* = p$.

Case: $\varkappa(p) \leq 0, \tau(p^*) > 0$. First note that the case $\varkappa(p) \leq 0, \tau(p^*) > 0$ is possible only if $p > 2$. It follows from (3.3) and $\tau(p^*) \leq \tau(p)$ since $\tau(\cdot)$ is decreasing. It implies $\tau(2) > 0$ and $\tau(r_l) > 0$ for any $l = 1, \dots, d$, since $r_l \leq p^*$.

Set $\varpi_\varepsilon = \varepsilon^2|\ln(\varepsilon)|$, and choose

$$A = \mathbf{c}_4 L_\beta^{1/(2\tau(2))}\varpi_\varepsilon^{(1-1/\omega)/(2\tau(2))}, \qquad m = 4,$$
$$\sigma_l = L_l^{-1/\beta_l}L_\beta^{(r_l-2)/(2\beta_l r_l\tau(2))}\varpi_\varepsilon^{\tau(r_l)/(2\beta_l\tau(2))}.$$

We remark first that

$$\sigma_l \to 0, \qquad \varepsilon \to 0 \qquad \forall l = 1, \dots, d,$$

and, therefore, $\sigma_l \leq b/2$ for all $\varepsilon > 0$ small enough.

Next, $|\mathcal{M}|/m = 4^{-1}L_\beta^{1/(\tau(2))}\varpi_\varepsilon^{-1/(2\beta\tau(2))} \geq 4^{-1}L_0^{1/(\beta\tau(2))}\varpi_\varepsilon^{-1/(2\beta\tau(2))}$, and hence for any ε small enough,

$$[\ln_2(|\mathcal{M}|/m - 1) - 2] \geq Q_3|\ln(\varepsilon)|,$$

where Q_3 is independent of ε and $\vec{L}$. This yields that (6.7) and (6.8) will be fulfilled if

$$\mathbf{c}_4^2 \le (2C_3)^{-1} Q_3, \qquad \mathbf{c}_4 4^{1/r_l} \le C_1^{-1}.$$

Choosing $\mathbf{c}_4$ small enough we satisfy the latter restrictions. Thus it remains to compute ρ_ε. We get from (6.9)

$$\begin{aligned} \rho_\varepsilon &= C_2 2^{1/p} \mathbf{c}_4 L_\beta^{(1-2/p)/(2\tau(2))} \varpi_\varepsilon^{\tau(p)/(2\tau(2))} \\ &=: Q_4 \big(L_\beta^{(1-2/p)/(\tau(p))} \varepsilon^2 |\ln(\varepsilon)|\big)^{(1-1/\omega+1/(\beta p))/(2-2/\omega+1/\beta)}. \end{aligned} \tag{6.12}$$

Case: $\varkappa(p) > 0$. Choose

$$\begin{aligned} &A = \mathbf{c}_6 (L_\beta \varepsilon^2)^{\beta/(2\beta+1)}, \qquad m = 9^{-1} L_\beta (L_\beta \varepsilon^2)^{-\beta/(2\beta+1)}, \\ &\sigma_l = L_l^{-1/\beta_l} (L_\beta \varepsilon^2)^{\beta/(\beta_l(2\beta+1))}. \end{aligned}$$

We remark that $|\mathcal{M}|/m = 9$ and $m \to \infty$, $\varepsilon \to 0$; hence $m > 4$. Moreover

$$\sigma_l \to 0, \qquad \varepsilon \to 0 \qquad \forall l = 1, \dots, d,$$

and therefore, $\sigma_l \le b/2$ for all $\varepsilon > 0$ small enough.

We obviously get that (6.7) and (6.8) will be fulfilled if

$$\mathbf{c}_6^2 \le (2C_3)^{-1}, \qquad \mathbf{c}_6 \le (9C_1)^{-1}.$$

Choosing $\mathbf{c}_6$ small enough we satisfy the latter restrictions. Finally, we get from (6.9),

$$\rho_\varepsilon = C_2 \mathbf{c}_4 18^{-1/p} (L_\beta \varepsilon^2)^{\beta/(2\beta+1)}.$$

7. Proof of Theorem 3.

Later on $\mathbf{c}_i$, $i = 1, 2, \dots,$ denote numerical constants independent of $\vec{L}$. Moreover without further mentioning we will assume that all quantities whose definitions involve the kernel $\mathcal{K}$ are defined with $\mathcal{K} = w_\ell$.

7.1. *Preliminary facts. Embedding of Nikol'skii classes.* For any $\vec{\beta} \in (0, \infty)^d$, $\vec{r} \in [1, \infty]^d$ and $s \ge 1$, define

$$\begin{aligned} &\boldsymbol{\gamma}_j(s) = \frac{\beta_j \tau(s)}{\tau(r_j)}, \qquad j = 1, \dots, d, \\ &\vec{\boldsymbol{\gamma}}(s) = (\boldsymbol{\gamma}_1(s) \wedge \beta_1, \dots, \boldsymbol{\gamma}_d(s) \wedge \beta_d); \end{aligned} \tag{7.1}$$

$$r^*(s) = \Big[\max_{j=1,\dots,d} r_j\Big] \vee s, \qquad \vec{r}(s) = (r_1 \vee s, \dots, r_d \vee s). \tag{7.2}$$

LEMMA 5. *For any $s \ge 1$, provided $\tau(r^*(s)) > 0$,*

$$\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \subseteq \mathbb{N}_{\vec{r}(s),d}(\vec{\boldsymbol{\gamma}}(s), \mathbf{c}\vec{L}),$$

where the constant $\mathbf{c} > 0$ is independent of $\vec{L}$, $\vec{r}$ and $\vec{\beta}$.

The statement of the lemma is a generalization of the embedding theorem for anisotropic Nikol'skii classes $\bar{\mathbb{N}}_{\vec{r},d}(\vec{\beta},\vec{L})$. Indeed, if $r^*(s)=s$ the assertion of the lemma can be found in Nikol'skiĭ (1977), Section 6.9. The proof of this lemma as well as that of Lemma 6 below is postponed to the Appendix.

Define $\mathcal{J}_\pm=\{j=1,\dots,d: r_j\neq\infty\}$, $p_\pm=[\sup_{j\in\mathcal{J}_\pm} r_j]\vee p$, and introduce

$$q_j=\begin{cases} p_\pm, & j\in\mathcal{J}_\pm,\\ \infty, & j\neq\mathcal{J}_\pm,\end{cases}\qquad \gamma_j=\begin{cases}\boldsymbol{\gamma}_j(p_\pm), & j\in\mathcal{J}_\pm,\\ \beta_j, & j\neq\mathcal{J}_\pm.\end{cases}\tag{7.3}$$

Note that $p^*\geq p_\pm$, and therefore if $\tau(p^*)>0$, we have in view of Lemma 5 with $s=p_\pm$

$$\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})\subseteq\mathbb{N}_{\vec{q},d}(\vec{\gamma},\mathbf{c}\vec{L}).\tag{7.4}$$

LEMMA 6. *Let $f\in\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{M})$, and let $\ell>\max_{j=1,\dots,d}\beta_j$. Then for any $\vec{\mathrm{h}}\in\mathfrak{S}_d^{\mathrm{const}}$,*

$$\begin{aligned}\|b_{\vec{\mathrm{h}},j}(\cdot,f)\|_{\mathbf{r},\mathbb{R}^d}&\leq(2b+1)^d\|w_\ell\|_{1,\mathbb{R}^d}\big(1-e^{-\beta_j}\big)^{-1}M_j\mathrm{h}_j^{\beta_j}\\ &\qquad\forall\mathbf{r}\in[1,r_j],\ j=1,\dots,d.\end{aligned}\tag{7.5}$$

Moreover, if $\tau(p^*)>0$, then for any $p\geq1$,

$$\begin{aligned}\|b_{\vec{\mathrm{h}},j}(\cdot,f)\|_{q_j,\mathbb{R}^d}&\leq(2b+1)^d\|w_\ell\|_{1,\mathbb{R}^d}\big(1-e^{-\gamma_j}\big)^{-1}M_j\mathrm{h}_j^{\gamma_j}\\ &\qquad\forall j=1,\dots,d,\end{aligned}\tag{7.6}$$

where $\vec{\gamma}$ and $\vec{q}$ are defined in (7.3).

7.2. *Preliminary facts. Maximal operator.* Let $\lambda:\mathbb{R}^m\to\mathbb{R}$, $m\geq1$ be a locally integrable function. We define the strong maximal function $M[\lambda]$ of λ by the formula

$$M[\lambda](x):=\sup_{\mathbb{K}_m}\frac{1}{\nu_m(\mathbb{K}_m)}\int_{\mathbb{K}_m}\lambda(t)\nu_m(\mathrm{d}t),\qquad x\in\mathbb{R}^m,\tag{7.7}$$

where the supremum is taken over all possible hyper-rectangles $\mathbb{K}_m$ in $\mathbb{R}^m$ with sides parallel to the coordinate axes, containing point x. It is worth noting that the *Hardy–Littlewood maximal function* is defined by (7.7) with the supremum taken over all cubes with sides parallel to the coordinate axes, centered at x.

It is well known that the strong maximal operator $\lambda\mapsto M[\lambda]$ is of the strong (r,r)-type for all $1<r\leq\infty$; that is, if $\lambda\in\mathbb{L}_r(\mathbb{R}^m)$, then $M[\lambda]\in\mathbb{L}_r(\mathbb{R}^m)$, and for any $r>1$, there exists a constant $\bar{C}(r)$ depending on r only such that

$$\|M[\lambda]\|_{r,\mathbb{R}^d}\leq\bar{C}(r)\|\lambda\|_{r,\mathbb{R}^d}.\tag{7.8}$$

Using the notation from Section 5.4, to any $J \subseteq \{1, \dots, d\} \cup \varnothing$ and locally integrable function $\lambda : \mathbb{R}^d \to \mathbb{R}_+$, we associate the operator

$$M_J[\lambda](x) = \sup_{\mathbb{K}_{|\bar{J}|}} \frac{1}{\nu_{|\bar{J}|}(\mathbb{K}_{|\bar{J}|})} \int_{\mathbb{K}_{|\bar{J}|}} \lambda\big(t + \mathbf{E}[J][x - t]\big)\nu_{|\bar{J}|}(\mathrm{d}t_{\bar{J}}),$$

where the supremum is taken over all hyper-rectangles in $\mathbb{R}^{|\bar{J}|}$ with center $x_{\bar{J}} = (x_j, j \in \bar{J})$ and with sides parallel to the axis.

As we see, $M[\lambda]$ is the strong maximal operator applied to the function obtained by λ by fixing by coordinates whose indices belong to J. It is obvious that $M_{\varnothing}[\lambda] \equiv M[\lambda]$ and $M_{\{1,\dots,d\}}[\lambda] \equiv \lambda$.

The following result is the direct consequence of (7.8) and Fubuni's theorem. For any $r > 1$, there exists $\mathbf{C}_r$ such that for any $d \ge 1$, λ, $J \subseteq \{1, \dots, d\} \cup \varnothing$ and $\mathrm{y} \in (0, \infty]$,

$$\|M_J[\lambda]\|_{r,(-\mathrm{y},\mathrm{y})^d} \le \mathbf{C}_r \|\lambda\|_{r,\mathcal{T}_J(\mathrm{y})}, \tag{7.9}$$

where we have denoted $\mathcal{T}_J(\mathrm{y}) = (-\mathrm{y}, \mathrm{y})^{|J|} \times \mathbb{R}^{|\bar{J}|}$. Note also that $\mathbf{C}_\infty = 1$.

7.3. *Preliminary facts*. *Key proposition*. The result presented in Proposition 2 below is the milestone for the proof of Theorem 3. For any $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$, define

$$\varphi := \varphi_\varepsilon(\vartheta, p) = \begin{cases} \big(L_\beta \varepsilon^2\big)^{\beta/(2\beta+1)}, & \varkappa(p) > 0; \\ \big(L_\beta \varepsilon^2 |\ln(\varepsilon)|\big)^{\beta/(2\beta+1)}, & \varkappa(p) \le 0. \end{cases}$$

Special set of bandwidths. For any $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$, $m \in \mathbb{N}$ and any $j = 1, \dots, d$, set

$$\widetilde{\boldsymbol{\eta}}_j(m) = e^{-2}\big(L_j^{-1}\varphi\big)^{1/\beta_j} e^{2dm(1/\beta_j - \omega(2+1/\beta)/(\beta_j r_j))}. \tag{7.10}$$

$$\hat{\boldsymbol{\eta}}_j(m) = e^{-2}\big(L_j^{-1}\varphi\big)^{1/\gamma_j} e^{2dm(1/\gamma_j - \upsilon(2+1/\gamma)/(\gamma_j q_j))} \left[\frac{L_\gamma \varphi^{1/\beta}}{L_\beta \varphi^{1/\gamma}}\right]^{\upsilon/(\gamma_j q_j)}, \tag{7.11}$$

where γ_j, q_j are defined in (7.3), and γ, υ and L_γ are given by

$$\frac{1}{\gamma} := \sum_{j=1}^d \frac{1}{\gamma_j}, \qquad \frac{1}{\upsilon} := \sum_{j=1}^d \frac{1}{\gamma_j q_j}, \qquad L_\gamma := \prod_{j=1}^d L_j^{1/\gamma_j}. \tag{7.12}$$

Introduce the integer $\widehat{\mathbf{m}} = \widehat{\mathbf{m}}(\vartheta, p)$, $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$, satisfying

$$\begin{aligned} &e^{-2d}\big[(L_\gamma/L_\beta)^{1/(1/\gamma - 1/\beta)}\varphi^{-1}\big]^{1/(2\beta\omega\tau(2))} \\ &\qquad \le e^{2d\widehat{\mathbf{m}}} \le \big[(L_\gamma/L_\beta)^{1/(1/\gamma - 1/\beta)}\varphi^{-1}\big]^{1/(2\beta\omega\tau(2))}. \end{aligned}$$

Later on, $\widehat{\mathbf{m}}$ will be used only if $\varkappa(p) < 0$ and $\tau(p^*) > 0$. Note that in this case $\widehat{\mathbf{m}} \ge 1$ for all $\varepsilon > 0$ small enough since $\tau(2) > 0$; see, for example, the proof of Theorem 2.

Introduce also the integer $\widetilde{\mathbf{m}} = \widetilde{\mathbf{m}}(\vartheta, p)$, $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$ as follows:

Case $\varkappa(p) > 0, \varkappa(p^*) \geq 0$: $\widetilde{\mathbf{m}} = +\infty$.

Case $\varkappa(p) > 0, \varkappa(p^*) < 0$: $e^{-2d}(\mathfrak{h}_\varepsilon^{-\ell} L_0^{-1}\varphi)^{p^*/(\varkappa(p^*))} \leq e^{2d\widetilde{\mathbf{m}}} \leq (\mathfrak{h}_\varepsilon^{-\ell} L_0^{-1}\varphi)^{p^*/(\varkappa(p^*))}$.

Case $\varkappa(p) \leq 0, \tau(p^*) \leq 0$: $e^{-2d}(L_0^{-1}\varphi)^{p^*/(\varkappa(p^*))} \leq e^{2d\widetilde{\mathbf{m}}} \leq (L_0^{-1}\varphi)^{p^*/(\varkappa(p^*))}$.

Case $\varkappa(p) \leq 0, \tau(p^*) > 0$: $\widetilde{\mathbf{m}} = \widehat{\mathbf{m}} + 1$ if $p^* = p$; $\widetilde{\mathbf{m}} = \widehat{\mathbf{m}} + \overline{\mathbf{m}}$ if $p^* > p$, where

$$e^{-2d}\varphi^{-(1+(1/\gamma-1/\beta)\upsilon(1/p-1/p^*))/((2+1/\gamma)\upsilon(1/p-1/p^*))}$$
$$\leq e^{2d\overline{\mathbf{m}}} \leq \varphi^{-(1+(1/\gamma-1/\beta)\upsilon(1/p-1/p^*))/((2+1/\gamma)\upsilon(1/p-1/p^*))}, \qquad p^* > p.$$

Some remarks are in order. First we note that $\widetilde{\mathbf{m}} \geq 1$ for all $\varepsilon > 0$ small enough. Indeed, $\varphi \to 0$, $\varepsilon \to 0$, and $\varkappa(p) \leq 0$ implies $\varkappa(p^*) < 0$ if $p^* > p$. Moreover, since $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$ the case $\varkappa(p) \leq 0, \tau(p^*) \leq 0$ is possible only if $p^* > p$ that, in its turn, implies $\varkappa(p^*) < 0$.

For any $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$ and any $0 \leq m \leq \widetilde{\mathbf{m}}$, introduce

$$\bar{\boldsymbol{\eta}}_j(m) = \begin{cases} \widetilde{\boldsymbol{\eta}}_j(m)1_{\{m \leq \widehat{\mathbf{m}}\}} + \widehat{\boldsymbol{\eta}}_j(m)_{1\{m > \widehat{\mathbf{m}}\}}, & \varkappa(p) \leq 0, \tau\big(p^*\big) > 0; \\ \widetilde{\boldsymbol{\eta}}_j(m), & \text{otherwise}, \end{cases} \tag{7.13}$$

and define $\vec{\boldsymbol{\eta}}(m) = (\boldsymbol{\eta}_1(m), \dots, \boldsymbol{\eta}_d(m))$ as follows.

For any $m \in \mathbb{N}$ set $\boldsymbol{\eta}_j(m) = \mathfrak{h}_{s_j(m)} \in \mathfrak{H}$, where $\mathbf{s}(m) = (s_1(m), \dots, s_d(m)) \in \mathbb{N}^d$ is given by

$$s_j(m) = \min\big\{s \in \mathbb{N} : \mathfrak{h}_s \leq \bar{\boldsymbol{\eta}}_j(m)\big\}. \tag{7.14}$$

Introduce finally the set of bandwidths $\mathfrak{H}_\varepsilon(\vartheta, p) = \{\vec{\boldsymbol{\eta}}(m), m = 0, \dots, \widetilde{\mathbf{m}}\}$.

LEMMA 7. *For any* $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$ *and any* $\varepsilon > 0$ *small enough, one has*

$$\mathfrak{H}_\varepsilon(\vartheta, p) \subset \begin{cases} \mathfrak{H}^d(\mathfrak{h}_\varepsilon), & \varkappa(p) > 0; \\ \mathfrak{H}^d, & \textit{otherwise}. \end{cases}$$

Moreover, $\mathbf{s}(m) \neq \mathbf{s}(n), \forall m \neq n, m, n = 0, \dots, \widetilde{\mathbf{m}}$.

Result formulation. For any $\vec{h} \in \mathfrak{S}_d$ and any $x \in \mathbb{R}^d$, put

$$\Phi_\varepsilon\big(V_{\vec{h}}(x)\big) = \begin{cases} V_{\vec{h}}^{-1/2}(x), & \varkappa(p) > 0; \\ \big[V_{\vec{h}}^{-1}(x)\big|\ln\big(\varepsilon V_{\vec{h}}(x)\big)\big|\big]^{1/2}, & \varkappa(p) \leq 0. \end{cases}$$

For any $g : \mathbb{R}^d \to \mathbb{R}$ and any $\vec{h} \in \mathfrak{S}_d^{\mathrm{const}}$, introduce

$$B_{\vec{h}}^*(x, g) = \sup_{\eta \in \mathfrak{S}_d^{\mathrm{const}}} B_{\vec{h}, \vec{\eta}}(x, g) + B_{\vec{h}}(x, g);$$
$$b_{\vec{h}}^*(x, g) = \sup_{J \in \mathfrak{J}} \sup_{j=1,\dots,d} M_J[b_{\vec{h},j}](x).$$

Here $B_{\vec{h},\vec{\eta}}$ and $B_{\vec{h}}$ are defined in (2.6), and $b_{\vec{h},j}$ is defined in (2.9), where f is replaced by g.

Let $\mathbb{C}_{\mathbb{K}}(\mathbb{R}^d)$ denote the set of continuous functions on $\mathbb{R}^d$ compactly supported on $\mathbb{K} = (-b-1, b+1)^d$, and let $\mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \vec{L}) = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \cap \mathbb{C}_{\mathbb{K}}(\mathbb{R}^d)$. Note that $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \subset \mathbb{C}(\mathbb{R}^d)$ if $\omega > 1$ in view of (7.4).

For any $\vec{\beta} \in (0,\infty)^d$ and $\vec{r} \in (1,\infty]^d$, set $\beta_* = \min_{j=1,\dots,d} \beta_j$, $\mathbf{C}(\vec{r}) = \max_{j=1,\dots,d} \mathbf{C}_{r_j}$, and define

$$\Upsilon_1 = 3d\big(1 \vee \|w_\ell\|_{\infty,\mathbb{R}^d}\big)^d, \qquad \Upsilon_2 = 4\Upsilon_1 \mathbf{C}(\vec{r})(2b+1)^d \|w_\ell\|_{1,\mathbb{R}^d}\big(1-e^{-\beta_*}\big)^{-1}.$$

PROPOSITION 2. *For any $\mathbf{a} \ge 1$, $\ell \in \mathbb{N}^*$, $L_0 > 0$, any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ and $\varepsilon > 0$, one can find $\mathfrak{S}^*_\varepsilon(\vartheta, p) = \{\vec{\mathbf{h}} : (-b,b)^d \to \mathfrak{H}_\varepsilon(\vartheta, p)\}$ such that for any $\varepsilon > 0$ small enough*:

(1) $\mathfrak{S}^*_\varepsilon(\vartheta, p) \subset \mathbb{H}_\varepsilon(3 + \sqrt{2b})$;

(2) *for any $g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L})$, there exists $\vec{\mathbf{h}}_g \in \mathfrak{S}^*_\varepsilon(\vartheta, p)$ such that*:

(i)

$$\begin{aligned} &B^*_{\vec{\mathbf{h}}_g}(x,g) + \mathbf{a}\Upsilon_2\varepsilon\Phi_\varepsilon\big(V_{\vec{\mathbf{h}}_g}(x)\big) \\ &\quad \le \inf_{\vec{h} \in \mathfrak{H}_\varepsilon(\vartheta,p)} \big[\Upsilon_1 b^*_{\vec{h}}(x,g) + \mathbf{a}\Upsilon_2\varepsilon\Phi_\varepsilon(V_{\vec{h}})\big] + \varepsilon \qquad \forall x \in (-b,b)^d; \end{aligned}$$

(ii) *if $\varkappa(p) > 0$ there exists $\mathfrak{r} \in \mathbb{N}^*_p$ such that $\varkappa(\frac{\mathfrak{r}p}{\mathfrak{r}-p}) > 0$ and $\mathfrak{r} \in \mathbb{N}^*_p(\vec{\mathbf{h}}_g, \mathcal{A}_\varepsilon)$.*

7.3.1. *Proof of Proposition 2.* We break the proof into several steps.

1^0. The condition $g \in \mathbb{C}_{\mathbb{K}}(\mathbb{R}^d)$ implies that g is uniformly continuous on $\mathbb{R}^d$, and therefore, for any $\varepsilon > 0$, there exists $\delta(\varepsilon)$ such that

$$|g(y) - g(y')| \le \varepsilon^2 \qquad \forall y, y' \in \mathbb{R}^d : |y - y'| \le \delta(\varepsilon). \tag{7.15}$$

Let $\mathrm{x}_{\mathbf{k},n}$, $\mathbf{k} \in \mathfrak{K}_n$, $n \in \mathbb{N}^*$ denote the center of the cube $\Delta^{(d)}_{\mathbf{k},n}$ defined in (3.6). Introduce for any $\vec{h} \in \mathfrak{S}^{\text{const}}_d$,

$$\widetilde{B}^*_{\vec{h}}(x,g) = \sum_{\mathbf{k} \in \mathfrak{K}_{\widetilde{n}}} B^*_{\vec{h}}(\mathrm{x}_{\mathbf{k},\widetilde{n}}, g) 1_{\Delta^{(d)}_{\mathbf{k},\widetilde{n}}}(x),$$

where $\widetilde{n}$ is chosen from the relation $2^{-\widetilde{n}} < \delta(\varepsilon) \le 2^{-\widetilde{n}+1}$.

Our first goal is to prove that

$$\sup_{\vec{h} \in \mathfrak{S}^{\text{const}}_d} \big\|B^*_{\vec{h}}(\cdot, g) - \widetilde{B}^*_{\vec{h}}(\cdot, g)\big\|_{\infty,\mathbb{R}^d} \le \mathbf{c}_1\varepsilon^2. \tag{7.16}$$

Indeed, for any $\vec{h} \in \mathfrak{S}^{\text{const}}_d$, since g is compactly supported on $\mathbb{K}$, one has

$$\big\|B^*_{\vec{h}}(\cdot, g) - \widetilde{B}^*_{\vec{h}}(\cdot, g)\big\|_{\infty,\mathbb{R}^d} = \sup_{\mathbf{k} \in \mathfrak{K}_{\widetilde{n}}} \sup_{x \in \Delta^{(d)}_{\mathbf{k},\widetilde{n}}} \big|B^*_{\vec{h}}(x,g) - B^*_{\vec{h}}(\mathrm{x}_{\mathbf{k},\widetilde{n}}, g)\big|. \tag{7.17}$$

In view of the definition of $B^*_{\vec{h}}(\cdot, g)$, we have for any $x \in \Delta^{(d)}_{\mathbf{k},\widetilde{n}}$,

$$\begin{aligned}\left|B^*_{\vec{h}}(x,g) - B^*_{\vec{h}}(\mathrm{x}_{\mathbf{k},\widetilde{n}}, g)\right| &\le 3 \sup_{\vec{\mathrm{h}} \in \mathfrak{S}_d^{\mathrm{const}}} \left|S_{\vec{\mathrm{h}}}(x,g) - S_{\vec{\mathrm{h}}}(\mathrm{x}_{\mathbf{k},\widetilde{n}}, g)\right| + \left|g(x) - g(\mathrm{x}_{\mathbf{k},\widetilde{n}})\right| \\ &\le 3 \sup_{\vec{\mathrm{h}} \in \mathfrak{S}_d^{\mathrm{const}}} \left|S_{\vec{\mathrm{h}}}(x,g) - S_{\vec{\mathrm{h}}}(\mathrm{x}_{\mathbf{k},\widetilde{n}}, g)\right| + \varepsilon^2.\end{aligned}$$

The last inequality follows from (7.15) and the definition of $\widetilde{n}$.

Recall that K is given in (3.5) and

$$S_{\vec{\mathrm{h}}}(x,g) = \int_{\mathbb{R}^d} K_{\vec{\mathrm{h}}}(t-x) g(t) \nu_d(\mathrm{d}t) = \int_{\mathbb{R}^d} K(u) g(x + u\vec{\mathrm{h}}) \nu_d(\mathrm{d}u).$$

Hence, for any $\vec{\mathrm{h}} \in \mathfrak{S}_d^{\mathrm{const}}$ and any $x \in \Delta^{(d)}_{\mathbf{k},\widetilde{n}}$,

$$\begin{aligned}\left|S_{\vec{\mathrm{h}}}(x,g) - S_{\vec{\mathrm{h}}}(\mathrm{x}_{\mathbf{k},\widetilde{n}}, g)\right| &\le \int_{\mathbb{R}^d} |K(u)| \left|g(x + u\vec{\mathrm{h}}) - g(\mathrm{x}_{\mathbf{k},\widetilde{n}} + u\vec{\mathrm{h}})\right| \nu_d(\mathrm{d}u) \\ &\le \|K\|_{1,\mathbb{R}^d} \varepsilon^2\end{aligned}$$

in view of (7.15) and the definition of $\widetilde{n}$.

Since the latter bound is independent of $\vec{\mathrm{h}}$, we obtain for any $\vec{h} \in \mathfrak{S}_d^{\mathrm{const}}$ and any $x \in \Delta^{(d)}_{\mathbf{k},\widetilde{n}}$,

$$\left|B^*_{\vec{h}}(x,g) - B^*_{\vec{h}}(\mathrm{x}_{\mathbf{k},\widetilde{n}}, g)\right| \le \left(1 + 3\|K\|_{1,\mathbb{R}^d}\right)\varepsilon^2.$$

Taking into account that the right-hand side of the latter inequality is independent of $\vec{h}$, $\mathbf{k}$ and x, we deduce (7.16) from (7.17).

One of the immediate consequences of (7.16) is that for any $x \in \mathbb{R}^d$,

$$\begin{aligned}&\Big|\inf_{\vec{h} \in \mathfrak{H}_\varepsilon(\vartheta,p)} \left[B^*_{\vec{h}}(x,g) + \mathbf{a}\Upsilon_2 \varepsilon \Phi_\varepsilon\left(V_{\vec{h}}(x)\right)\right] \\ &\qquad - \inf_{\vec{h} \in \mathfrak{H}_\varepsilon(\vartheta,p)} \left[\widetilde{B}^*_{\vec{h}}(x,g) + \mathbf{a}\Upsilon_2 \varepsilon \Phi_\varepsilon\left(V_{\vec{h}}(x)\right)\right]\Big| \le \mathbf{c}_1 \varepsilon^2.\end{aligned} \tag{7.18}$$

2^0. For any $x \in (-b, b)^d$, introduce

$$\vec{\mathbf{h}}_g(x) = \arg \inf_{\vec{\mathrm{h}} \in \mathfrak{H}_\varepsilon(\vartheta,p)} \left[\widetilde{B}^*_{\vec{\mathrm{h}}}(x,g) + \mathbf{a}\Upsilon_2 \varepsilon \Phi_\varepsilon\left(V_{\vec{h}}(x)\right)\right], \tag{7.19}$$

and define $\mathfrak{S}^*_\varepsilon(\vartheta, p) = \{\vec{\mathbf{h}}_g, g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L})\}$.

First, we deduce from (7.16) and (7.18) that for any $x \in (-b, b)^d$,

$$\begin{aligned}&B^*_{\vec{\mathbf{h}}_g}(x,g) + \mathbf{a}\Upsilon_2 \varepsilon \Phi_\varepsilon\left(V_{\vec{\mathbf{h}}_g}(x)\right) \\ &\qquad \le \inf_{\vec{h} \in \mathfrak{H}_\varepsilon(\vartheta,p)} \left[B^*_{\vec{h}}(x,g) + \mathbf{a}\Upsilon_2 \varepsilon \Phi_\varepsilon\left(V_{\vec{h}}(x)\right)\right] + 2\mathbf{c}_1 \varepsilon^2.\end{aligned} \tag{7.20}$$

Next, since $\widetilde{B}^*_{\mathrm{h}}(\cdot, g)$ is a piecewise constant on $\{\Delta^{(d)}_{\mathbf{k},\widetilde{n}} \cap (-b, b)^d, \mathbf{k} \in \mathfrak{K}_{\widetilde{n}}\}$, one has $\vec{\mathbf{h}}_g \in \mathfrak{S}^{\varepsilon}_{\widetilde{n}}$ and, hence we can assert that

$$\vec{\mathbf{h}}_g \in \bigcup_{n\in\mathbb{N}^*} \mathfrak{S}^{\varepsilon}_n \qquad \forall g \in \mathbb{C}_{\mathbb{K}}(\mathbb{R}^d). \tag{7.21}$$

Our goal now is to prove that for any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$, one can find $\varepsilon(\vartheta, p) > 0$ such that for any $\varepsilon < \varepsilon(\vartheta, p)$,

$$\vec{\mathbf{h}}_g \in \mathbb{H}_d\big(1/(2d), 3 + \sqrt{2b}, \mathcal{A}_\varepsilon\big) \qquad \forall g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \vec{L}). \tag{7.22}$$

3^0a. Note that the definition of the function w_ℓ together with the assumption $g \in \mathbb{C}_{\mathbb{K}}(\mathbb{R}^d)$ implies that $\sup_{x\in\mathbb{R}^d} |B^*_{\vec{h}}(x, g)| < \infty$, which implies, in view of (7.16), $\sup_{x\in\mathbb{R}^d} |\widetilde{B}^*_{\vec{h}}(x, g)| < \infty$.

Hence, $\mathbf{h}_{j,g}(x) < \infty$ for any $x \in (-b, b)^d$ and any $j = 1, \dots, d$, where $\mathbf{h}_{j,g}(\cdot)$ is jth coordinate of the vector-function $\vec{\mathbf{h}}_g$. It implies, in particular, that the infimum in (7.19) is achievable, and therefore, for any $x \in (-b, b)^d$,

$$\vec{\mathbf{h}}_g(x) \in \mathfrak{H}_\varepsilon(\vartheta, p) \qquad \forall g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L}). \tag{7.23}$$

By the same reasoning, $B^*_{\vec{h}}(\cdot, g)$ as well as $\widetilde{B}^*_{\vec{h}}(\cdot, g)$ are Borel functions, and since $\mathfrak{H}_\varepsilon(\vartheta, p)$ is countable, we assert in view of (7.23) that for any $\mathbf{s} \in \mathbb{N}^d$ such that $\vec{\mathfrak{h}}_{\mathbf{s}} := (\mathfrak{h}_{s_1}, \dots, \mathfrak{h}_{s_d}) \in \mathfrak{H}_\varepsilon(\vartheta, p)$,

$$\Lambda_{\mathbf{s}}[\vec{\mathbf{h}}_g] \in \mathfrak{B}\big(\mathbb{R}^d\big) \qquad \forall g \in \mathbb{C}_{\mathbb{K}}\big(\mathbb{R}^d\big). \tag{7.24}$$

This implies, in particular, that $\vec{\mathbf{h}}_g$ is Borel function.

3^0b. Taking into account that w_ℓ is compactly supported on $[-1/2, 1/2]^d$, we easily deduce from the assertions of Lemma 2 that for any $\vec{h}, \vec{\eta} \in \mathfrak{S}^{\text{const}}_d$,

$$B_{\vec{h},\vec{\eta}}(x, g) \le 2d\big(1 \vee \|w_\ell\|_{\infty,\mathbb{R}^d}\big)^d \sup_{J\in\mathfrak{J}} \sup_{j=1,\dots,d} M_J[b_{\vec{h},j}](x);$$

$$B_{\vec{h}}(x, g) \le d\big(1 \vee \|w_\ell\|_{\infty,\mathbb{R}^d}\big)^d \sup_{J\in\mathfrak{J}} \sup_{j=1,\dots,d} M_J[b_{\vec{h},j}](x).$$

Since the right-hand side of the first inequality is independent of $\vec{\eta}$, we obtain for any $\vec{h} \in \mathfrak{S}^{\text{const}}_d$,

$$B^*_{\vec{h}}(x, g) \le \Upsilon_1 \sup_{J\in\mathfrak{J}} \sup_{j=1,\dots,d} M_J[b_{\vec{h},j}](x) \qquad \forall x \in \mathbb{R}^d. \tag{7.25}$$

In particular, this yields, together with (7.20), that for any $x \in \mathbb{R}^d$,

$$\begin{aligned} &B^*_{\vec{\mathbf{h}}_g}(x, g) + \mathbf{a}\Upsilon_2\varepsilon\Phi_\varepsilon\big(V_{\vec{\mathbf{h}}_g}(x)\big) \\ &\qquad \le \inf_{\vec{h}\in\mathfrak{H}_\varepsilon(\vartheta,p)} \big[\Upsilon_1 b^*_{\vec{h}}(x, g) + \mathbf{a}\Upsilon_2\varepsilon\Phi_\varepsilon\big(V_{\vec{h}}(x)\big)\big] + 2\mathbf{c}_1\varepsilon^2. \end{aligned} \tag{7.26}$$

4^0. To get (7.22) let us first prove that for any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$, one can find $\varepsilon(\vartheta, p) > 0$ such that for any $\varepsilon < \varepsilon(\vartheta, p)$,

$$\vec{\mathbf{h}}_g \in \mathbb{H}_d\big(1/(2d), 3 + \sqrt{2b}\big) \qquad \forall g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L}). \tag{7.27}$$

For any $\mathbf{s} \in \mathbb{N}^*$ recall that $\vec{\mathfrak{h}}_{\mathbf{s}} = (\mathfrak{h}_{s_1}, \dots, \mathfrak{h}_{s_d})$ and $V_{\mathbf{s}} = \prod_{j=1}^d \mathfrak{h}_{s_j}$. Denote $\mathcal{S}^d = \{\mathbf{s}(m), m = 0, \dots, \widetilde{\mathbf{m}}\}$, and remark that $\Lambda_{\mathbf{s}}[\vec{\mathbf{h}}_g] := \{x \in (-b, b)^d : \vec{\mathbf{h}}_g(x) = \vec{\mathfrak{h}}_{\mathbf{s}}\} = \varnothing$ for any $\mathbf{s} \in \mathbb{N}^d, \mathbf{s} \neq \mathcal{S}^d$, in view of the definition of $\vec{\mathbf{h}}_g$.

Taking into account (7.24) we have for any $1 \le m \le \widetilde{\mathbf{m}}$ in view of the definition $\vec{\mathbf{h}}_g$ and the second assertion of Lemma 7,

$$\begin{aligned}
&\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]\\
&\quad\subseteq \big\{x \in (-b, b)^d :\\
&\qquad \widetilde{B}^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)}}(x, g) + \mathbf{a}\Upsilon_2\varepsilon\Phi_\varepsilon(V_{\mathbf{s}(m-1)}) \ge \widetilde{B}^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m)}}(x, g) + \mathbf{a}\Upsilon_2\varepsilon\Phi_\varepsilon(V_{\mathbf{s}(m)})\big\}\\
&\quad\subseteq \big\{x \in (-b, b)^d :\\
&\qquad \mathbf{c}_3\varepsilon^2 + B^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)}}(x, g) \ge \mathbf{a}\Upsilon_2\big[\varepsilon\Phi_\varepsilon(V_{\mathbf{s}(m)}) - \varepsilon\Phi_\varepsilon(V_{\mathbf{s}(m-1)})\big]\big\}.
\end{aligned}$$

To get the last inclusion we have taken into account (7.16). The definition of $\mathbf{s}(m)$ implies that

$$\begin{aligned}
&e^{-d}\prod_{j=1}^d \mathfrak{h}_{s_j(m)} \le \prod_{j=1}^d \bar{\boldsymbol{\eta}}_j(m) = e^{-2d}L_\beta^{-1}\varphi^{1/\beta}e^{-4dm} \le \prod_{j=1}^d \mathfrak{h}_{s_j(m)}\\
&\hspace{20em} \forall m = 0, \dots, \widetilde{\mathbf{m}}
\end{aligned} \tag{7.28}$$

and therefore, $V^{-1}_{\mathbf{s}(m-1)}V_{\mathbf{s}(m)} \le e^{-3d}$. It yields

$$\Phi_\varepsilon(V_{\mathbf{s}(m)}) - \Phi_\varepsilon(V_{\mathbf{s}(m-1)}) \ge 2^{-1}\Phi_\varepsilon(V_{\mathbf{s}})$$

for any $\varepsilon > 0$ small enough.

Putting $\mathbf{c}_2 = \mathbf{a}(2b + 1)\mathbf{C}(\vec{r})\|w\|_{1,\mathbb{R}^d}(1 - e^{-\beta_*})^{-1}$ and using (7.25), we have for any $\varepsilon > 0$, provided $\varepsilon < \mathbf{c}_1^{-1}\mathbf{c}_2$,

$$\begin{aligned}
\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g] &\subseteq \big\{x \in (-b, b)^d : \mathbf{c}_1\varepsilon^2 + B^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)}}(x, g) \ge 2^{-1}\mathbf{a}\Upsilon_2\varepsilon\Phi_\varepsilon(V_{\mathbf{s}})\big\}\\
&\subseteq \bigcup_{J \in \mathfrak{J}}\bigcup_{j=1}^d \big\{x \in (-b, b)^d : \varepsilon^{-1}\Phi_\varepsilon^{-1}(V_{\mathbf{s}})M_J[b_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)}, j}](x) > \mathbf{c}_2\big\}.
\end{aligned} \tag{7.29}$$

Here we have also used that $\Upsilon_1 \ge 1$ as well as $\Phi_\varepsilon(V_{\mathbf{s}}) > 1$ for any $\mathbf{s} \in \mathcal{S}^d$.

Introduce $\mathcal{J}_\infty = \{j = 1, \dots, d : r_j = \infty\}$, and recall that $\mathcal{J}_\pm = \{1, \dots, d\} \setminus \mathcal{J}_\infty$. In view of (7.9) and the bound (7.5) of Lemma 6 with $\mathbf{r} = \infty$ and $\vec{M} = \mathbf{a}\vec{L}$, we obtain for any $j \in \mathcal{J}_\infty$ and any $J \in \mathfrak{J}$,

$$\begin{aligned}
\|M_J[b_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)}, j}]\|_{\infty,\mathbb{R}^d} &\le \mathbf{c}_2 L_j \mathfrak{h}^{\beta_j}_{s_j(m-1)}\\
&\le \mathbf{c}_2 L_j \bar{\boldsymbol{\eta}}_j^{\beta_j}(m - 1) \le \mathbf{c}_2\varphi e^{2d(m-1)}.
\end{aligned} \tag{7.30}$$

Here we have used that $\bar{\boldsymbol{\eta}}_j^{\beta_j}(m-1) = e^{-2}(L_j^{-1}\varphi)^{1/\beta_j} e^{2d(m-1)}$ if $r_j = \infty$.

Set $\mu_\varepsilon = 1$ if $\varkappa(p) > 0$ and $\mu_\varepsilon = \sqrt{|\ln(\varepsilon)|}$ if $\varkappa(p) \le 0$. We obtain for any $m = 0, \dots, \widetilde{\mathbf{m}}$ in view of (7.28) and the definition of φ,

$$(\varepsilon\mu_\varepsilon)^{-1}\sqrt{V_{\mathbf{s}(m)}}\varphi e^{2d(m-1)} \le e^{-5d/2}.$$

Moreover, we obviously have that $\Phi(V_{\mathbf{s}}) \ge V_{\mathbf{s}}^{-1/2}\mu_\varepsilon$ for any $\mathbf{s} \in \mathbb{N}^d$. Thus we have

$$\varepsilon^{-1}\Phi_\varepsilon^{-1}(V_{\mathbf{s}(m)})\varphi e^{2d(m-1)} \le e^{-5d/2}, \tag{7.31}$$

and therefore, for any $j \in J_\infty$ and any $J \in \mathfrak{J}$,

$$\|M_J[b_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)},j}]\|_{\infty,\mathbb{R}^d} \le \mathbf{c}_2 e^{-5d/2} < \mathbf{c}_2.$$

This yields, together with (7.29),

$$\begin{aligned}\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g] \\ \subseteq \bigcup_{J\in\mathfrak{J}}\bigcup_{j\in\mathcal{J}_\pm}^{d}\{x \in (-b,b)^d : \varepsilon^{-1}\Phi_\varepsilon^{-1}(V_{\mathbf{s}})M_J[b_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)},j}](x) > \mathbf{c}_2\}.\end{aligned} \tag{7.32}$$

We remark also that if $J_\pm = \varnothing$, then only $\Lambda_{\mathbf{s}(0)}[\vec{\mathbf{h}}_g] \ne \varnothing$. Let us consider now separately two cases.

4^0a. Suppose that either $\varkappa(p) > 0$ or $\varkappa(p) \le 0, \tau(p^*) \le 0$, and recall that $\bar{\boldsymbol{\eta}}_j(m) = \widetilde{\boldsymbol{\eta}}(m)$, $j = 1, \dots, d$, for all values of m.

Applying the Markov inequality, we get for any $m = 1, \dots, \widetilde{\mathbf{m}}$, in view of (7.9) and the bound (7.5) of Lemma 6 with $\mathbf{r} = r_j$ and $\vec{M} = \mathbf{a}\vec{L}$,

$$\begin{aligned}2^{-d}\nu_d(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]) &\le \sum_{j\in\mathcal{J}_\pm}^{d}[\mathbf{c}_2\varepsilon\Phi_\varepsilon(V_{\mathbf{s}})]^{-r_j}\|b_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)},j}\|_{r_j,\mathbb{R}^d}^{r_j} \\ &\le \sum_{j\in\mathcal{J}_\pm}^{d}[\mathbf{c}_2\varepsilon\Phi_\varepsilon(V_{\mathbf{s}})]^{-r_j}(\mathbf{c}_2 L_j \mathfrak{h}_{s_j(m-1)}^{\beta_j})^{r_j} \\ &\le \sum_{j\in\mathcal{J}_\pm}^{d}[\varepsilon^{-1}\Phi_\varepsilon^{-1}(V_{\mathbf{s}})L_j\widetilde{\boldsymbol{\eta}}_j^{\beta_j}(m-1)]^{r_j} \\ &\le \sum_{j\in\mathcal{J}_\pm}^{d}[\varepsilon^{-1}\Phi_\varepsilon^{-1}(V_{\mathbf{s}})\varphi e^{2d(m-1)}]^{r_j}e^{-2d\omega(2+1/\beta)(m-1)}.\end{aligned} \tag{7.33}$$

Taking into account that $\omega \ge \beta$, we obtain in view of (7.31) that for any $\varepsilon < \mathbf{c}_1^{-1}\mathbf{c}_2$,

$$\begin{aligned}\nu_d(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]) \le d e^{-d/2}e^{-2d\omega(2+1/\beta)(m-1)} \le d e^{-d/2}e^{-2d(m-1)} \\ \forall m = 1, \dots, \widetilde{\mathbf{m}}.\end{aligned} \tag{7.34}$$

Remembering that $\Lambda_{\mathbf{s}}[\vec{\mathbf{h}}_g]=\varnothing$ for any $\mathbf{s}\notin\mathcal{S}_d$ and that $\nu_d(\Lambda_{\mathbf{s}(0)}[\vec{\mathbf{h}}_g])\le(2b)^d$, taking into account the second assertion of Lemma 7, we obtain, putting $\tau=(2d)^{-1}$ for any $\varepsilon<\mathbf{c}_1^{-1}\mathbf{c}_2$,

$$\begin{aligned}\sum_{\mathbf{s}\in\mathbb{N}^d}\nu_d^{\tau}\big(\Lambda_{\mathbf{s}}[\vec{\mathbf{h}}_g]\big)&=\sum_{m=1}^{\widetilde{\mathbf{m}}}\nu_d^{\tau}\big(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]\big)+(2b)^{d\tau}\le d^{1/(2d)}\big(1-e^{-1}\big)^{-1}+\sqrt{2b}\\&\le 2+\sqrt{2b}.\end{aligned}$$

Here we have used that $\sup_{d\ge1}d^{1/(2d)}(1-e^{-1})^{-1}<2$.

Thus we assert that (7.27) is established if either $\varkappa(p)>0$ or $\varkappa(p)\le 0,\tau(p^*)\le0$.

4^0b. Let now $\varkappa(p)\le0,\tau(p^*)>0$. Since $\bar{\boldsymbol{\eta}}_j(m)=\widetilde{\boldsymbol{\eta}}(m)$, $j=1,\dots,d$, if $m=0,\dots,\widehat{\mathbf{m}}$, (7.34) remains true for any $m=0,\dots,\widehat{\mathbf{m}}$. Similarly to (7.33) we obtain for any $m>\widehat{\mathbf{m}}$ in view of (7.9) and bound (7.6) of Lemma 6 with $\vec{M}=\mathbf{a}\vec{L}$,

$$\begin{aligned}2^{-d}\nu_d\big(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]\big)&\le\sum_{j\in\mathcal{J}_\pm}^{d}\big[\mathbf{c}_2\varepsilon\Phi_\varepsilon(V_{\mathbf{s}})\big]^{-q_j}\|b_{\vec{\mathfrak{h}}_{\mathbf{s}(m-1)},j}\|_{q_j,\mathbb{R}^d}^{q_j}\\&\le\sum_{j\in\mathcal{J}_\pm}^{d}\big[\mathbf{c}_2\varepsilon\Phi_\varepsilon(V_{\mathbf{s}})\big]^{-q_j}\big(\mathbf{c}_3\big(1-e^{-\gamma_j}\big)^{-1}L_j\mathfrak{h}_{s_j(m-1)}^{\gamma_j}\big)^{q_j}\\&\le\sum_{j\in\mathcal{J}_\pm}^{d}\big[\mathbf{c}_4\big(1-e^{-\gamma_j}\big)^{-1}\varepsilon^{-1}\Phi_\varepsilon^{-1}(V_{\mathbf{s}})L_j\widehat{\boldsymbol{\eta}}_j^{\gamma_j}(m-1)\big]^{q_j},\end{aligned}\tag{7.35}$$

where we have put $\mathbf{c}_3=(1-e^{\beta_*})\mathbf{c}_2$ and $\mathbf{c}_4=(1-e^{\beta_*})$.

Using (7.31) we get

$$\big[\varepsilon^{-1}\Phi_\varepsilon^{-1}(V_{\mathbf{s}})L_j\widehat{\boldsymbol{\eta}}_j^{\gamma_j}(m-1)\big]^{q_j}\le e^{-3dp_\pm/2}e^{-2d\upsilon(2+1/\gamma)(m-1)}\left[\frac{L_\gamma\varphi^{1/\beta}}{L_\beta\varphi^{1/\gamma}}\right]^{\upsilon}.$$

Moreover, the definition of $\widehat{\mathbf{m}}$ implies that

$$\begin{aligned}&e^{-2d\upsilon(2+1/\gamma)\widehat{\mathbf{m}}}\varphi^{(1/\beta-1/\gamma)\upsilon}\\&\quad\le e^d(L_\beta/L_\gamma)^{-\upsilon(2+1/\gamma)(2\beta\omega\tau(2)(1/\gamma-1/\beta))}\varphi^{\upsilon(2+1/\gamma)/(2\beta\omega\tau(2))-\upsilon(1/\gamma-1/\beta)}.\end{aligned}$$

Below we prove [see formula (A.20)] that $\upsilon(2+1/\gamma)-\omega(2+1/\beta)=2\beta\tau(2)\omega\upsilon(1/\gamma-1/\beta)$, and we obtain [recall that $\tau(2)>0$ in the considered case; see, e.g., the proof of Theorem 2]

$$\begin{aligned}&e^{-2d\upsilon(2+1/\gamma)\widehat{\mathbf{m}}}\varphi^{(1/\beta-1/\gamma)\upsilon}\\&\quad\le e^d(L_\beta/L_\gamma)^{-\upsilon(2+1/\gamma)/(2\beta\omega\tau(2)(1/\gamma-1/\beta))}\varphi^{(2+1/\beta)/(2\beta\tau(2))}\to0,\qquad\varepsilon\to0.\end{aligned}$$

The latter bound together with (7.35) yields that for any $m \geq \widehat{\mathbf{m}}+1$,

$$\begin{aligned}
&\nu_d\big(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]\big)\\
&\quad \leq e^{-3d/2}(L_\beta/L_\gamma)^{(2+1/\beta)/(2(\beta/\gamma-1)\tau(2))}\varphi^{(2+1/\beta)/(2\beta\tau(2))}\\
&\qquad \times e^{-2d\upsilon(2+1/\gamma)(m-\widehat{\mathbf{m}}-1)},
\end{aligned}\tag{7.36}$$

and therefore, putting $\tau=(2d)^{-1}$, we can assert that one can find $\varepsilon(\vartheta,p)>0$ such that for any $\varepsilon<\varepsilon(\vartheta,p)$,

$$\sum_{m=\widehat{\mathbf{m}}+1}^{\widetilde{\mathbf{m}}} \nu_d^\tau\big(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]\big) \leq 1.$$

This yields, together with (7.34), for all $\varepsilon<\min\{\mathbf{c}_1^{-1}\mathbf{c}_2,\varepsilon(\vartheta,p)\}$,

$$\sum_{\mathbf{s}\in\mathbb{N}^d} \nu_d^\tau\big(\Lambda_{\mathbf{s}}[\vec{\mathbf{h}}_g]\big) \leq \sum_{m=1}^{\widetilde{\mathbf{m}}} \nu_d^\tau\big(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]\big)+(2b)^{d\tau}+1 \leq 3+\sqrt{2b}.$$

Thus we assert that (7.27) is established in the case $\varkappa(p)\leq 0$, $\tau(p^*)>0$ as well.

5^0. To get (7.22) it remains to prove that for any $(\vartheta,p)\in\mathcal{P}^{\mathrm{consist}}$, one can find $\varepsilon(\vartheta,p)>0$ such that for any $\varepsilon<\varepsilon(\vartheta,p)$,

$$\vec{\mathbf{h}}_g \in \mathbb{B}(\mathcal{A}_\varepsilon) \qquad \forall g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L}).\tag{7.37}$$

The proof of (7.37) is mostly based on the choice of $\mathcal{A}_\varepsilon$ given in (2.2) which, in turn, guarantees (2.3). We will consider separately 2 cases.

5^0a. Let $\varkappa(p)>0$. Obviously we can find $\mathfrak{r}\in\mathbb{N}^*_p$ such that $\mathfrak{p}:=\frac{p\mathfrak{r}}{\mathfrak{r}-p}$ satisfies $\varkappa(\mathfrak{p})>0$, and we have in view of (7.28),

$$\|V_{\vec{\mathbf{h}}_g}^{-1/2}\|_{\mathfrak{p}}^{\mathfrak{p}} \leq e^{d\mathfrak{p}}L_\beta^{\mathfrak{p}/2}\varphi^{-\mathfrak{p}/(2\beta)}\Bigg[(2b)^d+\sum_{m=1}^{\mathbf{m}} e^{2\mathfrak{p}dm}\nu_d\big(\Lambda_{\mathbf{s}(m)}[\vec{\mathbf{h}}_g]\big)\Bigg].\tag{7.38}$$

Using the first bound established in (7.34), we obtain

$$\begin{aligned}
&\|V_{\vec{\mathbf{h}}_g}^{-1/2}\|_{\mathfrak{p}}^{\mathfrak{p}}\\
&\quad \leq e^{d\mathfrak{p}}L_\beta^{\mathfrak{p}/2}\varphi^{-\mathfrak{p}/(2\beta)}\Bigg[(2b)^d+d2^{-d/2}e^{-2d\omega(2+1/\beta)}\sum_{m=1}^{\mathbf{m}} e^{2(\mathfrak{p}-\omega(2+1/\beta))dm}\Bigg].
\end{aligned}\tag{7.39}$$

Remembering that $\mathfrak{p}-\omega(2+1/\beta)=:-\varkappa(\mathfrak{p})<0$ and that $\varphi^{\mathfrak{p}/(2\beta)}\mathcal{A}_\varepsilon\to\infty$, we assert that there exists $\varepsilon(\theta,p)>0$ such that $\|V_{\vec{\mathbf{h}}_g}^{-1/2}\|_{\mathfrak{p}}^{\mathfrak{p}}\leq\mathcal{A}_\varepsilon$ for any $\varepsilon<\varepsilon(\theta,p)$.

Thus (7.37) is proved if $\varkappa(p)>0$. Moreover, since the right-hand side of inequality (7.39) as well as the choice of $\mathfrak{r}$ are independent of g, we can assert that for all $\varepsilon<\varepsilon(\theta,p)$,

$$\mathfrak{r}\in\mathbb{N}^*_p(\vec{\mathbf{h}}_g,\mathcal{A}_\varepsilon) \qquad \forall g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L}),$$

and assertion 2(ii) of the proposition follows.

5^0b. In all other cases the set $\mathfrak{H}_\varepsilon(\vartheta, p)$ is finite, and we obviously have in view of (7.28),

$$\|V_{\vec{\mathbf{h}}_g}^{-1/2}\|_t \le (2b)^d e^d L_\beta^{1/2} \varphi^{-1/(2\beta)} e^{2d\tilde{\mathbf{m}}} \qquad \forall t \ge 1. \tag{7.40}$$

This, together with the definitions of $\tilde{\mathbf{m}}$ and φ, imply that the right-hand side of the latter inequality increases to infinity polynomially in ε^{-1}. Thus there exists $\varepsilon(\theta, p) > 0$ such that $\|V_{\vec{\mathbf{h}}_g}^{-1/2}\|_{\mathfrak{p}}^{\mathfrak{p}} \le \mathcal{A}_\varepsilon$ for any $\varepsilon < \varepsilon(\theta, p)$, and (7.37) follows.

6^0. Thus (7.22) follows from (7.27) and (7.37), and it yields, together with (7.21), that $\mathfrak{S}_\varepsilon^*(\vartheta, p) \subset \mathbb{H}_\varepsilon(R)$ for all $\varepsilon > 0$ small enough, and therefore, the first assertion of the proposition is proved. We note that assertion 2(i) of the proposition follows from (7.26) for any $\varepsilon > 0$ such that $2\mathbf{c}_1\varepsilon \le 1$. Recall, at last, that in view of (7.23) any $\vec{\mathbf{h}} \in \mathfrak{S}_\varepsilon^*(\vartheta, p)$ takes values in $\mathfrak{H}_\varepsilon(\vartheta, p)$.

7.4. *Proof of the theorem. Case $p \in (1, \infty)$.* We will need some technical results presented in Lemmas 8 and 9 whose proofs are postponed to the Appendix.

Recall that the quantity $\mathcal{B}_{\vec{h}}^{(p)}(\cdot)$ is defined in (2.7) with K given in (3.5). Furthermore, $\mathbf{a} = \|K\|_{1,\mathbb{R}^d} = \|w_\ell\|_{1,\mathbb{R}}^d$.

LEMMA 8. *For any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ and any $\mathbb{H} \subseteq \mathfrak{S}_d$,*

$$\sup_{f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})} \inf_{\vec{h} \in \mathbb{H}} \big[\mathcal{B}_{\vec{h}}^{(p)}(f) + \varepsilon \Psi_{\varepsilon,p}(\vec{h})\big] \le \sup_{g \in \mathbb{N}_{\vec{r},d}^*(\vec{\beta}, \mathbf{a}\vec{L})} \inf_{\vec{h} \in \mathbb{H}} \big[\mathcal{B}_{\vec{h}}^{(p)}(g) + \varepsilon \Psi_{\varepsilon,p}(\vec{h})\big].$$

For any $x \in (-b, b)^d$ and any $g \in \mathbb{N}_{\vec{r},d}^*(\vec{\beta}, \mathbf{a}\vec{L})$, define

$$U_{\vartheta,p}(x, g) = \inf_{\vec{h} \in \mathfrak{H}_\varepsilon(\vartheta, p)} \big[b_{\vec{h}}^*(x, g) + \varpi_\varepsilon V_{\vec{h}}^{-1/2}\big],$$

where $\varpi_\varepsilon = \varepsilon$ if $\varkappa(p) > 0$ and $\varpi_\varepsilon = \varepsilon\sqrt{|\ln(\varepsilon)|}$ if $\varkappa(p) \le 0$.

LEMMA 9. *For any $(\vartheta, p) \in \mathcal{P}$, provided $p^* > p$ and any $\varepsilon > 0$ small enough,*

$$\sup_{g \in \mathbb{N}_{\vec{r},d}^*(\vec{\beta}, \mathbf{a}\vec{L})} \|U_{\vartheta,p}(\cdot, g)\|_{p^*} \le \Upsilon_3 L^*,$$

where, recall, $L^* = \min_{j:r_j=p^*} L_j$ and $\Upsilon_3 = \mathbf{a}d2^d\mathbf{C}_{p^*}(\mathbf{C}_{p^*}\|w_\ell\|_{\infty,\mathbb{R}^d} + 1) + 1$.

Let $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ be fixed. Later on $R = 3 + \sqrt{2b}$, and without further mentioning we will assume that $\varepsilon > 0$ is sufficiently small in order to provide the results of Proposition 2, Lemmas 7 and 9. Set also

$$V_p(\vec{L}) = (L_\gamma/L_\beta)^{(p-\omega(2+1/\beta))/(2p\beta\omega\tau(2)(1/\gamma-1/\beta))} L_\beta^{\tau(p)/(2\tau(2))},$$
$$p < \infty;\ V_\infty(\vec{L}) = L_\gamma.$$

7.4.1. *Proof of the theorem. Preliminaries.* We deduce from Theorem 1 and Lemma 8

$$\mathcal{R} := \mathbf{c}_5^{-1} \sup_{f \in \mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})} \mathcal{R}_\varepsilon^{(p)}[\hat{f}_{\vec{\mathbf{h}}}^{(R)}; f]$$

$$\leq \sup_{g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \inf_{\vec{h} \in \mathbb{H}_\varepsilon(R)} \{\mathcal{B}_{\vec{h}}^{(p)}(g) + \varepsilon \Psi_{\varepsilon,p}(\vec{h})\} + \varepsilon.$$

In view of the first assertion of Proposition 2, $\mathfrak{S}^*_\varepsilon(\theta, p) \subset \mathbb{H}_\varepsilon(R)$.

$$\mathcal{R} \leq \sup_{g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \inf_{\vec{h} \in \mathfrak{S}^*_\varepsilon(\theta,p)} \{\mathcal{B}_{\vec{h}}^{(p)}(g) + \varepsilon \Psi_{\varepsilon,p}(\vec{h})\} + \varepsilon. \tag{7.41}$$

Note also the following obvious inequality: for any $p \geq 1$ and any $g : \mathbb{R}^d \to \mathbb{R}$,

$$\sup_{\vec{\eta} \in \mathfrak{S}_d} \|B_{\vec{h},\vec{\eta}}(\cdot, g)\|_p \leq \Big\| \sup_{\eta \in \mathfrak{S}_d^{\mathrm{const}}} B_{\vec{h},\vec{\eta}}(\cdot, g)\Big\|_p \qquad \forall \vec{h} \in \mathfrak{S}_d.$$

This yields, in particular, for any $\vec{h} \in \mathfrak{S}_d$ and any $g : \mathbb{R}^d \to \mathbb{R}$,

$$\mathcal{B}_{\vec{h}}^{(p)}(g) \leq 2\|B^*_{\vec{h}}(\cdot, g)\|_p. \tag{7.42}$$

Combining (7.41) and (7.42) we get

$$\mathcal{R} \leq \sup_{g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \{2\|B^*_{\vec{\mathbf{h}}_g}(\cdot, g)\|_p + \varepsilon \Psi_{\varepsilon,p}(\vec{\mathbf{h}}_g)\} + \varepsilon, \tag{7.43}$$

where $\vec{\mathbf{h}}_g$ satisfies the second assertion of Proposition 2. Consider separately two cases.

Case $\varkappa(p) > 0$. Recall that $\vec{\mathbf{h}}_g(x)$ takes values in $\mathfrak{H}_\varepsilon(\vartheta, p)$ for any $g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L})$ and $x \in (-b, b)^d$. Additionally, $\mathfrak{H}_\varepsilon(\vartheta, p) \subset \mathfrak{H}^d(\mathfrak{h}_\varepsilon)$ in view of the first assertion of Lemma 7 since $\varkappa(p) > 0$.

This implies $\vec{\mathbf{h}}_g \in \mathfrak{S}_d(\mathfrak{h}_\varepsilon)$, and we can assert that

$$\Psi_{\varepsilon,p}(\vec{\mathbf{h}}_g) \leq \inf_{r \in \mathbb{N}^*_p(\vec{\mathbf{h}}_g,\mathcal{A}_\varepsilon)} C_2(r)\|V_{\vec{\mathbf{h}}_g}^{-1/2}\|_{rp/(r-p)} \qquad \forall g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L}).$$

Applying assertion 2(ii) of Proposition 2, we can state that for some $\mathfrak{r}$, provided $\varkappa(\frac{\mathfrak{r}p}{\mathfrak{r}-p}) > 0$,

$$\Psi_{\varepsilon,p}(\vec{\mathbf{h}}_g) \leq C_2(\mathfrak{r})\|V_{\vec{\mathbf{h}}}^{-1/2}\|_{\mathfrak{r}p/(\mathfrak{r}-p)} \qquad \forall g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L}).$$

Denote $\mathfrak{p} = \frac{\mathfrak{r}p}{\mathfrak{r}-p}$. Since $\mathfrak{p} > p$ in view of Hölder's inequality, $\|B^*_{\vec{\mathbf{h}}_g}(\cdot, g)\|_p \leq (2b)^d\|B^*_{\vec{\mathbf{h}}_g}(\cdot, g)\|_{\mathfrak{p}}$, and we deduce from (7.43) (remembering that we consider here the norms of positive functions) that

$$
\begin{aligned}
\mathcal{R} &\le \mathbf{c}_6 \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \big\{\big\|B^*_{\vec{\mathbf{h}}_g}(\cdot,g)\big\|_{\mathfrak{p}} + \varepsilon\big\|V^{-1/2}_{\vec{\mathbf{h}}_g}\big\|_{\mathfrak{p}}\big\} + \varepsilon \\
&\le \mathbf{c}_7 \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \big\|B^*_{\vec{\mathbf{h}}_g}(\cdot,g) + \mathbf{a}\Upsilon_2\varepsilon V^{-1/2}_{\vec{\mathbf{h}}_g}(\cdot)\big\|_{\mathfrak{p}} + \varepsilon.
\end{aligned}
\tag{7.44}
$$

Applying assertion 2(i) of Proposition 2 we obtain

$$
\mathcal{R} \le \mathbf{c}_8 \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \Big\| \inf_{\vec{h}\in\mathfrak{H}_\varepsilon(\vartheta,p)} \big[b^*_{\vec{h}}(\cdot,g) + \varepsilon\Phi_\varepsilon(V_{\vec{h}})\big]\Big\|_{\mathfrak{p}} + \varepsilon.
\tag{7.45}
$$

Case $\varkappa(p)\le 0$. Since $\Psi_{\varepsilon,p}(\vec{h}) \le (C_1\|\sqrt{|\ln(\varepsilon V_{\vec{h}})|}V^{-1/2}_{\vec{h}}\|_p)$ for any $\vec{h}\in\mathbb{B}(\mathcal{A}_\varepsilon)$, we deduce from, similarly to (7.44),

$$
\mathcal{R} \le \mathbf{c}_9 \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \Big\|B^*_{\vec{\mathbf{h}}_g}(\cdot,g) + \mathbf{a}\Upsilon_2\varepsilon\sqrt{\big|\ln\big(\varepsilon V_{\vec{\mathbf{h}}_g}(\cdot)\big)\big|}V^{-1/2}_{\vec{\mathbf{h}}_g}(\cdot)\Big\|_p + \varepsilon.
$$

Applying the first assertion 2(i) of Proposition 2 we have

$$
\mathcal{R} \le \mathbf{c}_{10} \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \Big\| \inf_{\vec{h}\in\mathfrak{H}_\varepsilon(\vartheta,p)} \big[b^*_{\vec{h}}(\cdot,g) + \varepsilon\Phi_\varepsilon(V_{\vec{h}})\big]\Big\|_p + \varepsilon.
$$

This together with (7.45) allows us to assert that

$$
\mathcal{R} \le \mathbf{c}_{11} \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \Big\| \inf_{\vec{h}\in\mathfrak{H}_\varepsilon(\vartheta,p)} \big[b^*_{\vec{h}}(\cdot,g) + \varepsilon\Phi_\varepsilon(V_{\vec{h}})\big]\Big\|_{\mathbf{p}} + \varepsilon,
\tag{7.46}
$$

where we have denoted $\mathbf{p}=\mathfrak{p}$ if $\varkappa(p)>0$ and $\mathbf{p}=p$ if $\varkappa(p)\le 0$.

The definition of $\widetilde{\mathbf{m}}$ allows us to assert that if $\varkappa(p)\le 0$,

$$
\Phi_\varepsilon(V_{\vec{h}}) \le \mathbf{c}_{12}\sqrt{|\ln(\varepsilon)|}V^{-1/2}_{\vec{h}} \qquad \forall\vec{h}\in\mathfrak{H}_\varepsilon(\vartheta,p).
$$

Hence we get from (7.46),

$$
\begin{aligned}
\mathcal{R} &\le \mathbf{c}_{13} \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \Big\| \inf_{\vec{h}\in\mathfrak{H}_\varepsilon(\vartheta,p)} \big[b^*_{\vec{h}}(\cdot,g) + \varpi_\varepsilon V^{-1/2}_{\vec{h}}\big]\Big\|_{\mathbf{p}} + \varepsilon \\
&= \mathbf{c}_{13} \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \big\|U_{\vartheta,p}(\cdot,g)\big\|_{\mathbf{p}} + \varepsilon =: \mathbf{c}_{13} \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \mathcal{R}_{\mathbf{p}}(g) + \varepsilon,
\end{aligned}
\tag{7.47}
$$

where we recall that $\varpi_\varepsilon=\varepsilon$ if $\varkappa(p)>0$ and $\varpi_\varepsilon=\varepsilon\sqrt{|\ln(\varepsilon)|}$ if $\varkappa(p)\le 0$.

7.4.2. *Proof of the theorem. Slicing.* For any $g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L})$ we have

$$\mathcal{R}^{\mathbf{p}}_{\mathbf{p}}(g) \le (2b)^d (Q\varphi)^{\mathbf{p}} + \sum_{m=0}^{\widetilde{\mathbf{m}}} \big(Qe^{2d(m+1)}\varphi\big)^{\mathbf{p}} \nu_d(\Gamma_m)$$

$$+ \int_{\Gamma_{\mathbf{m}}} \big|U_{\vartheta,p}(x,g)\big|^{\mathbf{p}} \nu_d(\mathrm{d}x) \tag{7.48}$$

$$=: (2b)^d (Q\varphi)^{\mathbf{p}} + \big(Qe^{2d}\varphi\big)^{\mathbf{p}} \sum_{m=1}^{\widetilde{\mathbf{m}}} e^{2dm\mathbf{p}} \nu_d(\Gamma_m) + T_{\widetilde{\mathbf{m}}}.$$

Here we have put $\Gamma_m = \{x \in (b,b)^d : U_{\vartheta,p}(x,g) \ge Qe^{2dm}\varphi\}$ and $Q = 2\mathbf{c}_2 + e^d$, where we recall that $\mathbf{c}_2 = \mathbf{a}(2b+1)\mathbf{C}(\vec{r})\|w\|_{1,\mathbb{R}^d}(1-e^{-\beta_*})^{-1}$. Moreover, if $\widetilde{\mathbf{m}} = \infty$, we set $T_\infty = 0$.

We have, in view of the definition of $U_{\vartheta,p}$,

$$\Gamma_m \subset \big\{x \in (b,b)^d : b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m)}}(\cdot,g) + \varpi_\varepsilon V^{-1/2}_{\mathbf{s}(m)} \ge Qe^{2dm}\varphi\big\}.$$

Recall that in view of (7.28) $e^{-d}V_{\mathbf{s}(m)} \le e^{-2d}L_\beta^{-1}\varphi^{1/\beta}e^{-4dm} \le V_{\mathbf{s}(m)}$ for any $m = 0,\dots,\widetilde{\mathbf{m}}$.

Hence $\varpi_\varepsilon V^{-1/2}_{\mathbf{s}(m)} \le e^d\varphi e^{2dm}$, and we get

$$\Gamma_m \subset \big\{x \in (b,b)^d : b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m)}}(\cdot,g) \ge 2\mathbf{c}_2 e^{2dm}\varphi\big\} =: \Gamma^*_m. \tag{7.49}$$

Note also that $\varpi_\varepsilon V^{-1/2}_{\mathbf{s}(m)} \le e^d\varphi e^{2dm} < e^d\mathbf{c}_2 e^{2dm}\varphi < e^d b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m)}}(x,g)$ for any $x \in \Gamma^*_m$, which yields

$$\big|U_{\vartheta,p}(x,g)\big| \le \big(e^d+1\big) b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(m)}}(x,g) \qquad \forall x \in \Gamma^*_m. \tag{7.50}$$

The latter inequality allows us to bound from above $T_{\widetilde{\mathbf{m}}}$ if $\widetilde{\mathbf{m}} < \infty$. Indeed, in view of (7.50),

$$T_{\widetilde{\mathbf{m}}} \le \big(e^d+1\big)^{\mathbf{p}} \big\|b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(\widetilde{\mathbf{m}})}}(\cdot,g)\big\|^{\mathbf{p}}_{\mathbf{p}}. \tag{7.51}$$

Another bound can be obtained in the case $p^* > \mathbf{p}$. Applying Hölder's inequality, the assertion of Lemma 9 and (7.4.2),

$$T_{\widetilde{\mathbf{m}}} \le \big(L^*\Upsilon_3\big)^{\mathbf{p}} \big[\nu_d\big(\Gamma^*_{\widetilde{\mathbf{m}}}\big)\big]^{1-\mathbf{p}/p^*}. \tag{7.52}$$

The definition of $b^*_{\vec{\mathfrak{h}}_{\mathbf{s(m)}}}(\cdot,g)$ implies that for any $m = 1,\dots,\widetilde{\mathbf{m}}$,

$$\Gamma^*_m \subset \bigcup_{J\in\mathfrak{J}} \bigcup_{j=1}^{d} \big\{x \in (b,b)^d : M_J\big[b_{\vec{\mathfrak{h}}_{\mathbf{s}(m),j}}\big](x) \ge 2\mathbf{c}_2 e^{2dm}\varphi\big\}.$$

Since in view of (7.30) $\|M_J[b_{\vec{\mathfrak{h}}_{\mathbf{s}(m)},j}]\|_{\infty,\mathbb{R}^d} \le \mathbf{c}_2\varphi e^{2dm}$ for any $j \in \mathcal{J}_\infty$ and any $J \in \mathfrak{J}$, we obtain

$$\Gamma_m^* \subset \bigcup_{J\in\mathfrak{J}} \bigcup_{j\in\mathcal{J}_\pm} \{x \in (b,b)^d : M_J[b_{\vec{\mathfrak{h}}_{\mathbf{s}(m)},j}](x) \ge \mathbf{c}_2 e^{2dm}\varphi\} \qquad \forall m = 1,\dots,\widetilde{\mathbf{m}}. \tag{7.53}$$

If either $\varkappa(p) > 0$ or $\varkappa(p) \le 0$, $\tau(p^*) \le 0$, the following bound is true:

$$\nu_d(\Gamma_m^*) \le \mathbf{c}_{14} e^{-2dm\omega(2+1/\beta)} \qquad \forall m = 1,\dots,\widetilde{\mathbf{m}}. \tag{7.54}$$

Indeed, applying the Markov inequality, we get for any $m = 1,\dots,\widetilde{\mathbf{m}}$, in view of (7.9) and the bound (7.5) of Lemma 6 with $\mathbf{r} = r_j$ and $\vec{M} = \mathbf{a}\vec{L}$,

$$\begin{aligned} 2^{-d}\nu_d(\Gamma_m^*) &\le \sum_{j\in\mathcal{J}_\pm}^{d} [\mathbf{c}_2 e^{2dm}\varphi]^{-r_j} \|b_{\vec{\mathfrak{h}}_{\mathbf{s}(m)},j}\|_{r_j,\mathbb{R}^d}^{r_j} \le \sum_{j\in\mathcal{J}_\pm}^{d} [e^{2dm}\varphi]^{-r_j} (L_j \mathfrak{h}_{s_j(m)}^{\beta_j})^{r_j} \\ &\le \sum_{j\in\mathcal{J}_\pm}^{d} [e^{-2dm}\varphi^{-1} L_j \widetilde{\boldsymbol{\eta}}_j^{\beta_j}(m)]^{r_j} \le d e^{-2dm\omega(2+1/\beta)}. \end{aligned}$$

If $\varkappa(p) \le 0$, $\tau(p^*) > 0$, we have for any $\widehat{\mathbf{m}} < m \le \widetilde{\mathbf{m}}$,

$$\nu_d(\Gamma_m^*) \le \mathbf{c}_{16}(L_\gamma/L_\beta)^{\upsilon} \varphi^{(1/\beta-1/\gamma)\upsilon} e^{-2dm\upsilon(2+1/\gamma)}. \tag{7.55}$$

Indeed, we obtain for any $m > \widehat{\mathbf{m}}$ in view of (7.9) and the bound (7.6) of Lemma 6 with $\vec{M} = \mathbf{a}\vec{L}$,

$$\begin{aligned} 2^{-d}\nu_d(\Gamma_m^*) &\le \sum_{j\in\mathcal{J}_\pm}^{d} [\mathbf{c}_2 e^{2dm}\varphi]^{-q_j} \|b_{\vec{\mathfrak{h}}_{\mathbf{s}(m)},j}\|_{q_j,\mathbb{R}^d}^{q_j} \\ &\le \mathbf{c}_{15} \sum_{j\in\mathcal{J}_\pm}^{d} [e^{2dm}\varphi]^{-q_j} (L_j \mathfrak{h}_{s_j(m)}^{\gamma_j})^{q_j} \\ &\le \mathbf{c}_{15} \sum_{j\in\mathcal{J}_\pm}^{d} [e^{-2dm}\varphi^{-1} L_j \widehat{\boldsymbol{\eta}}_j^{\gamma_j}(m)]^{q_j} \\ &\le \mathbf{c}_{16}(L_\gamma/L_\beta)^{\upsilon} \varphi^{(1/\beta-1/\gamma)\upsilon} e^{-2dm\upsilon(2+1/\gamma)}. \end{aligned}$$

7.4.3. *Proof of the theorem. Derivation of rates.* We will now proceed differently depending to which zone the pair $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$ belongs.

Dense zone: $\varkappa(p) > 0$. *Case* $\varkappa(p^*) \ge 0$. Recall that $\widetilde{\mathbf{m}} = \infty$ in this case, and therefore $T_{\widetilde{\mathbf{m}}} = 0$. Moreover $\mathbf{p} = \mathfrak{p}$. We get from (7.48) and (7.54),

$$\mathcal{R}_{\mathfrak{p}}^{\mathfrak{p}}(g) \le \mathbf{c}_{17}\varphi^{\mathfrak{p}} \sum_{m=0}^{\infty} e^{2dm(\mathfrak{p}-\omega(2+1/\beta))} \le \mathbf{c}_{18}\varphi^{\mathfrak{p}} = \mathbf{c}_{18}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}\mathfrak{p}} \tag{7.56}$$

since $\mathfrak{p} - \omega(2 + 1/\beta) = -\varkappa(\mathfrak{p}) < 0$ in view of the definition of $\mathfrak{p}$. Taking into account that the right-hand side of the latter inequality is independent of g, we obtain in view of (7.47),

$$\mathcal{R} \le \mathbf{c}_{19}(\boldsymbol{\delta}_\varepsilon)^{\mathfrak{a}}. \tag{7.57}$$

Dense zone: $\varkappa(p) > 0$. *Case* $\varkappa(p^*) < 0$. Note first that $\varkappa(p^*) < 0$ and $\varkappa(\mathfrak{p}) > 0$ implies $\mathfrak{p} < p^*$ since $\varkappa(\cdot)$ is decreasing. We get in view of (7.52) and (7.54),

$$\begin{aligned} T_{\widetilde{\mathbf{m}}} &\le \mathbf{c}_{20}(L^*)^{\mathfrak{p}} e^{-2d\widetilde{\mathbf{m}}\omega(2+1/\beta)(1-\mathfrak{p}/p^*)} \\ &\le \mathbf{c}_{21}(L^*)^{\mathfrak{p}} (\mathfrak{h}_\varepsilon)^{\ell p^*/(\varkappa(p^*))} \varphi^{-\omega(2+1/\beta)(1-\mathfrak{p}/p^*)/(\varkappa(p^*)/p^*)}. \end{aligned} \tag{7.58}$$

Noting that $\mathfrak{p} + \frac{\omega(2+1/\beta)(1-\mathfrak{p}/p^*)}{\varkappa(p^*)/p^*} = \frac{p^*\varkappa(\mathfrak{p})}{\varkappa(p^*)} < 0$ and taking into account (2.3), we obtain that

$$\varphi^{-\mathfrak{p}} (\mathfrak{h}_\varepsilon)^{\ell p^*/(\varkappa(p^*))} T_{\widetilde{\mathbf{m}}} \to 0, \qquad \varepsilon \to 0.$$

Since (7.56) holds, we assert finally that (7.57) remains true if $\varkappa(p^*) < 0$ as well. Thus the theorem is proved in the case $\varkappa(p) > 0$.

New zone: $\varkappa(p) \le 0$, $\tau(p^*) \le 0$. Recall that $\mathbf{p} = p$ and necessarily $p^* > p$ since we consider $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$.

Noting that the first inequality in (7.58) remains true, we deduce from (7.48) and (7.54),

$$\begin{aligned} \mathcal{R}_p^p(g) &\le \mathbf{c}_{22}\varphi^p \sum_{m=0}^{\widetilde{\mathbf{m}}} e^{2dm(p-\omega(2+1/\beta))} \\ &\qquad + \mathbf{c}_{20}(L^*)^p e^{-2d\widetilde{\mathbf{m}}\omega(2+1/\beta)(1-p/p^*)}. \end{aligned} \tag{7.59}$$

If $\varkappa(p) < 0$, we have

$$\begin{aligned} \mathcal{R}_p^p(g) &\le \mathbf{c}_{23}\varphi^p e^{2d\widetilde{\mathbf{m}}(p-\omega(2+1/\beta))} + \mathbf{c}_{20}(L^*)^p e^{-2d\widetilde{\mathbf{m}}\omega(2+1/\beta)(1-p/p^*)} \\ &\le \mathbf{c}_{24}(1+L^*)^p \varphi^{(pp^*\omega(2+1/\beta)(1/p-1/p^*))/(p^*-\omega(2+1/\beta))} \\ &= \mathbf{c}_{24}(1+L^*)^p \boldsymbol{\delta}_\varepsilon^{(pp^*\omega(1/p-1/p^*))/(p^*-\omega(2+1/\beta))} \end{aligned} \tag{7.60}$$

in view of the definition of $\widetilde{\mathbf{m}}$. Taking into account that the right-hand side of the latter inequality is independent of g, we obtain in view of (7.47) that

$$\mathcal{R} \le \mathbf{c}_{19}(1+L^*)(\boldsymbol{\delta}_\varepsilon)^{\mathfrak{a}}. \tag{7.61}$$

If $\varkappa(p) = 0$, we deduce from (7.59) and the definition of $\widetilde{\mathbf{m}}$ that

$$\begin{aligned} \mathcal{R}_p^p(g) &\le \mathbf{c}_{23}\varphi^p \widetilde{\mathbf{m}} + (L^*)^p \varphi^{(pp^*\omega(2+1/\beta)(1/p-1/p^*))/(p^*-\omega(2+1/\beta))} \\ &\le \mathbf{c}_{24}\varphi^p |\ln(\varepsilon)| + (L^*)^p \varphi^p. \end{aligned}$$

Here we have used that $\frac{p^*\omega(2+1/\beta)(1/p-1/p^*)}{p^*-\omega(2+1/\beta)} = \frac{\beta}{2\beta+1}$ if $\varkappa(p)=0$. Thus we conclude

$$\mathcal{R} \le \mathbf{c}_{25}(\boldsymbol{\delta}_\varepsilon)^{\mathfrak{a}}\bigl|\ln(\varepsilon)\bigr|^{1/p}. \tag{7.62}$$

Thus the theorem is proved in the case $\varkappa(p) \le 0$.

Sparse zone: $\varkappa(p) \le 0$, $\tau(p^*) > 0$. If $p^* = p$, taking into account that $\widetilde{\mathbf{m}} = \widehat{\mathbf{m}} + 1$, we deduce from (7.48), (7.51), (7.54) and (7.55) that

$$\mathcal{R}_p^p(g) \le \mathbf{c}_{18}\varphi^p \sum_{m=0}^{\widehat{\mathbf{m}}} e^{2dm(p-\omega(2+1/\beta))} + \bigl(e^d+1\bigr)^p \bigl\|b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(\widehat{\mathbf{m}}+1)}}(\cdot, g)\bigr\|_p^p.$$

Using (7.51), (7.9) and the triangle inequality, we have

$$\bigl\|b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(\widehat{\mathbf{m}}+1)}}(\cdot, g)\bigr\|_p \le \sum_{J\in\mathfrak{J}}\sum_{j=1}^{d}\bigl\|M_J[b_{\vec{\mathfrak{h}}_{\mathbf{s}(\widehat{\mathbf{m}}+1)},j}\bigr\|_p \le \mathbf{c}_{26}\sum_{j=1}^{d}\bigl\|b_{\vec{\mathfrak{h}}_{\mathbf{s}(\widehat{\mathbf{m}}+1)},j}\bigr\|_{p,\mathbb{R}^d}.$$

Note that $p^* = p$ implies $p_\pm = p$, and therefore, $q_j = p$ for any $j = 1,\dots,d$, where we recall that q_j are given in (7.3). Hence we obtain, using the bound (7.6) of Lemma 6 with $\vec{M} = \mathbf{a}\vec{L}$,

$$\begin{aligned}
&\bigl\|b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(\widehat{\mathbf{m}}+1)}}(\cdot, g)\bigr\|_p \\
&\quad\le \mathbf{c}_{27}\sum_{j=1}^{d} L_j \mathfrak{h}^{\gamma_j}_{s_j(\widehat{\mathbf{m}}+1)} \le \mathbf{c}_{27}\sum_{j=1}^{d} L_j \widehat{\boldsymbol{\eta}}_j^{\gamma_j}(\widehat{\mathbf{m}}+1) \\
&\quad\le \mathbf{c}_{28}(L_\gamma/L_\beta)^{\upsilon/p}\varphi^{1+(1/\beta-1/\gamma)(\upsilon/p)}e^{2d\widehat{\mathbf{m}}(1-(\upsilon/p)(2+1/\gamma))} \\
&\quad\le \mathbf{c}_{28}(L_\gamma/L_\beta)^{(\omega(2+1/\beta)-p)/(2p\beta\omega\tau(2)(1/\gamma-1/\beta))}\varphi^{((2+1/\beta)\tau(p))/(2\tau(2))} \\
&\quad= \mathbf{c}_{28}(L_\gamma/L_\beta)^{(\omega(2+1/\beta)-p)/(2p\beta\omega\tau(2)(1/\gamma-1/\beta))} \\
&\qquad\times L_\beta^{\tau(p)/(2\tau(2))}\bigl(\varepsilon^2\bigl|\ln(\varepsilon)\bigr|\bigr)^{\tau(p)/(2\tau(2))} \\
&\quad= \mathbf{c}_{28}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}}.
\end{aligned} \tag{7.63}$$

We get, in view of the definition of $\widehat{\mathbf{m}}$,

$$\begin{aligned}
\varphi^p \sum_{m=0}^{\widehat{\mathbf{m}}} e^{2dm(p-\omega(2+1/\beta))} &\le \mathbf{c}_{29}\varphi^p e^{2d\widehat{\mathbf{m}}(p-\omega(2+1/\beta))} \\
&\le \mathbf{c}_{30}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}p}, \qquad \varkappa(p) < 0;
\end{aligned} \tag{7.64}$$

$$\begin{aligned}
\varphi^p \sum_{m=0}^{\widehat{\mathbf{m}}} e^{2dm(p-\omega(2+1/\beta))} &\le \mathbf{c}_{29}\varphi^p(\widehat{\mathbf{m}}+1) \\
&\le \mathbf{c}_{29}\varphi^p\bigl|\ln(\varepsilon)\bigr|, \qquad \varkappa(p) = 0.
\end{aligned} \tag{7.65}$$

Therefore, if $\varkappa(p) < 0$,

$$\mathcal{R}_p^p(g) \le \mathbf{c}_{31}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}p}. \tag{7.66}$$

If $\varkappa(p) = 0$, we can easily check that $\frac{\tau(p)}{2\tau(2)} = \frac{\beta}{2\beta+1}$, which yields, in view of the definition of $\widehat{\mathbf{m}}$,

$$\mathcal{R}_p^p(g) \le \mathbf{c}_{32}\varphi^p\big|\ln(\varepsilon)\big| = \mathbf{c}_{33}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}p}\big|\ln(\varepsilon)\big|. \tag{7.67}$$

Taking into account that the right-hand sides in (7.66) and (7.67) are independent of g, we obtain, in view of (7.47),

$$\begin{aligned} &\mathcal{R} \le \mathbf{c}_{34}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}}, \qquad \varkappa(p) < 0; \\ &\mathcal{R} \le \mathbf{c}_{29}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}}\big|\ln(\varepsilon)\big|, \qquad \varkappa(p) = 0. \end{aligned} \tag{7.68}$$

This completes the proof of the theorem in the case $\varkappa(p) \le 0$, $\tau(p^*) > 0$, $p^* = p$.

If $p^* > p$, we deduce from (7.48), (7.54) and (7.55) that

$$\begin{aligned} \mathcal{R}_p^p(g) &\le \mathbf{c}_{18}\varphi^p \sum_{m=0}^{\widehat{\mathbf{m}}} e^{2dm(p-\omega(2+1/\beta))} \\ &\quad + \mathbf{c}_{16}(L_\gamma/L_\beta)^\upsilon \varphi^{p+(1/\beta-1/\gamma)\upsilon} \sum_{m=\widehat{\mathbf{m}}+1}^{\widetilde{\mathbf{m}}} e^{2dm(p-\upsilon(2+1/\gamma))} + T_{\widetilde{\mathbf{m}}} \\ &\le \mathbf{c}_{35}\Bigg[\varphi^p \sum_{m=0}^{\widehat{\mathbf{m}}} e^{2dm(p-\omega(2+1/\beta))} \\ &\quad + (L_\gamma/L_\beta)^\upsilon \varphi^{p+(1/\beta-1/\gamma)\upsilon} e^{2d\widehat{\mathbf{m}}(p-\upsilon(2+1/\gamma))}\Bigg] + T_{\widetilde{\mathbf{m}}}. \end{aligned}$$

Here we have used that $p \le p_\pm < \upsilon(2+1/\gamma) < 0$ in view of (A.18).

Using (A.20) and the definition of $\widehat{\mathbf{m}}$ we compute that

$$\begin{aligned} &(L_\gamma/L_\beta)^\upsilon \varphi^{p+(1/\beta-1/\gamma)\upsilon} e^{2d\widehat{\mathbf{m}}(p-\upsilon(2+1/\gamma))} \\ &\quad \le \mathbf{c}_{36}(L_\gamma/L_\beta)^{(\omega(2+1/\beta)-p)/(2\beta\omega\tau(2)(1/\gamma-1/\beta))} \varphi^{p(2+1/\beta)\tau(p)/(2\tau(2))} \\ &\quad = \mathbf{c}_{36}V_p^p(\vec{L})\big(\varepsilon^2\big|\ln(\varepsilon)\big|\big)^{p\tau(p)/(2\tau(2))} = \mathbf{c}_{36}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}p}. \end{aligned}$$

This yields, together with (7.64) and (7.65),

$$\begin{aligned} &\mathcal{R}_p^p(g) \le \mathbf{c}_{37}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}p} + T_{\widetilde{\mathbf{m}}}, \qquad \varkappa(p) < 0; \\ &\mathcal{R}_p^p(g) \le \mathbf{c}_{29}\boldsymbol{\delta}_\varepsilon^{\mathfrak{a}p}\big|\ln(\varepsilon)\big| + T_{\widetilde{\mathbf{m}}}, \qquad \varkappa(p) = 0. \end{aligned}$$

Using (7.52) and (7.55),

$$\begin{aligned} T_{\widetilde{\mathbf{m}}} &\le \mathbf{c}_{38}A\varphi^{(1/\beta-1/\gamma)\upsilon(1-p/p^*)} e^{-2d\widetilde{\mathbf{m}}\upsilon(2+1/\gamma)(1-p/p^*)} \\ &\le \mathbf{c}_{38}A\varphi^{(1/\beta-1/\gamma)\upsilon(1-p/p^*)} e^{-2d\overline{\mathbf{m}}\upsilon(2+1/\gamma)(1-p/p^*)} = \mathbf{c}_{38}A\varphi^p \end{aligned}$$

in view of the definition of $\widetilde{\mathbf{m}}$ and $\overline{\mathbf{m}}$. Here $A = A(L_\beta, L_\gamma, L^*)$ can be easily computed.

Thus we can assert that (7.68) holds in the case $p^* > p$ as well, which completes the proof of the theorem.

7.5. *Proof of the theorem. Case* $p \in \{1, \infty\}$. Note that $p_\pm = \infty$ if $p = \infty$ and therefore $\vec{\gamma} = \vec{\boldsymbol{\gamma}}(\infty)$. The proof of the theorem in this case is the straightforward consequence of Corollary 1.

Introduce the vectors $\vec{u} = (u_1, \dots, u_d)$ and $\vec{v} = (v_1, \dots, v_d)$ as follows:

$$u_j = L_j^{-1/\beta_j}\left(L_\beta \varepsilon^2\right)^{\beta/(\beta_j(2\beta+1))}, \qquad v_j = L_j^{-1/\gamma_j}\left(L_\gamma \varepsilon^2 |\ln(\varepsilon)|\right)^{\gamma/(\gamma_j(2\gamma+1))}.$$

It is obvious that both vectors belong to $\mathbb{H}_\varepsilon^{\text{const}}$, and without loss of generality we can assume that $\vec{u}, \vec{v} \in \mathfrak{H}^d$. Moreover,

$$\varepsilon \Psi_{\varepsilon,\infty}^{(\text{const})}(\vec{v}) \le \mathbf{c}_{38} \varepsilon \sqrt{|\ln(\varepsilon)|} V_{\vec{v}}^{-1/2} = \mathbf{c}_{38}\left(L_\gamma \varepsilon^2 |\ln(\varepsilon)|\right)^{\gamma/(2\gamma+1)}.$$

Note that $1/\gamma = \sum_{j=1}^d \frac{\tau(r_j)}{\beta_j \tau(\infty)} = \frac{1}{\beta \tau(\infty)} \Rightarrow \frac{\gamma}{2\gamma+1} = \frac{\tau(\infty)}{2\tau(2)}$ and therefore,

$$\varepsilon \Psi_{\varepsilon,\infty}^{(\text{const})}(\vec{v}) = \mathbf{c}_{38} \boldsymbol{\delta}^{\mathfrak{a}}.$$

Additionally, we easily compute

$$\varepsilon \Psi_{\varepsilon,1}^{(\text{const})}(\vec{v}) \le \mathbf{c}_{39} \varepsilon V_{\vec{v}}^{-1/2} = \mathbf{c}_{38}\left(L_\beta \varepsilon^2\right)^{\beta/(2\beta+1)} = \mathbf{c}_{39} \boldsymbol{\delta}^{\mathfrak{a}}.$$

Applying the assertions of Lemma 6, we obtain

$$\sum_{j=1}^d \|b_{\vec{u},j}\|_1 \le \mathbf{c}_{40} \sum_{j=1}^d L_j u_j^{\beta_j} = \mathbf{c}_{40} \boldsymbol{\delta}^{\mathfrak{a}}; \qquad \sum_{j=1}^d \|b_{\vec{v},j}\|_\infty \le \mathbf{c}_{41} \sum_{j=1}^d L_j v_j^{\gamma_j} = \mathbf{c}_{41} \boldsymbol{\delta}^{\mathfrak{a}}.$$

The assertion of the theorem follows now from Corollary 1.

APPENDIX

A.1. Proof of the assertion (ii) of Lemma 1. For any given $\mathbf{s} \in \mathbb{N}^d$ and any $\vec{\mathrm{h}} \in \mathfrak{S}_d$, define

$$\Lambda_{s_j}[\mathrm{h}_j] = \left\{x \in (-b, b)^d : \mathrm{h}_j(x) = s_j\right\}, \qquad j = 1, \dots, d.$$

Then $\Lambda_{\mathbf{s}}[\vec{\mathrm{h}}] = \bigcap_{j=1}^d \Lambda_{s_j}[\mathrm{h}_j]$ and we get putting $\mathbf{s}_j = (s_1, \dots, s_{j-1}, s_{j+1}, \dots, d)$

$$\nu_d\left(\Lambda_{s_j}[\mathrm{h}_j]\right) = \sum_{\mathbf{s}_j \in \mathbb{N}^{d-1}} \nu_d\left(\Lambda_{\mathbf{s}}[\vec{\mathrm{h}}]\right), \qquad j = 1, \dots, d.$$

This yields, for any $\alpha \in (0, 1)$, that we have for any $\vec{\mathrm{h}} \in \mathfrak{S}_d$,

$$\sum_{s_j=1}^\infty \nu_d^{\alpha/d}\left(\Lambda_{s_j}[\mathrm{h}_j]\right) \le \sum_{\mathbf{s} \in \mathbb{N}^d} \nu_d^{\alpha/d}\left(\Lambda_{\mathbf{s}}[\vec{\mathrm{h}}]\right).$$

Since obviously $\vec{h} \vee \vec{\eta} \in \mathfrak{S}_d$ for any $\vec{h}, \vec{\eta} \in \mathfrak{S}_d$, we have

$$\sum_{s_j=1}^{\infty} \nu_d^{\alpha/d}\big(\Lambda_{s_j}[h_j \vee \eta_j]\big) \le \sum_{s_j=1}^{\infty} \big\{\nu_d^{\alpha/d}\big(\Lambda_{s_j}[h_j]\big) + \nu_d^{\alpha/d}\big(\Lambda_{s_j}[\eta_j]\big)\big\}.$$

Hence, for any $\alpha \in (0, 1)$ and any $\vec{h}, \vec{\eta} \in \mathbb{H}_d(d^{-1}\alpha, 2^{-1}L^{1/d})$, $L > 0$, we get

$$\sum_{s_j=1}^{\infty} \nu_d^{\alpha/d}\big(\Lambda_{s_j}[h_j \vee \eta_j]\big) \le \sum_{\mathbf{s}\in\mathbb{N}^d} \nu_d^{\alpha/d}\big(\Lambda_{\mathbf{s}}[\vec{h}]\big) + \sum_{\mathbf{s}\in\mathbb{N}^d} \nu_d^{\alpha/d}\big(\Lambda_{\mathbf{s}}[\vec{\eta}]\big) \le L^{1/d}. \tag{A.1}$$

Note that $\Lambda_{\mathbf{s}}[\vec{\mathrm{h}}] = \bigcap_{j=1}^{d} \Lambda_{s_j}[\mathrm{h}_j]$ implies that for any $\vec{\mathrm{h}} \in \mathfrak{S}_d$ and $\alpha \in (0, 1)$,

$$\nu_d^{\alpha}\big(\Lambda_{\mathbf{s}}[\vec{\mathrm{h}}]\big) \le \prod_{j=1}^{d} \nu_d^{\alpha/d}\big(\Lambda_{s_j}[\mathrm{h}_j]\big). \tag{A.2}$$

Therefore, we deduce from (A.1) and (A.2) that

$$\begin{aligned}\sum_{\mathbf{s}\in\mathbb{N}^d} \nu_d^{\alpha}\big(\Lambda_{\mathbf{s}}[\vec{h} \vee \vec{\eta}]\big) &\le \sum_{\mathbf{s}\in\mathbb{N}^d} \prod_{j=1}^{d} \nu_d^{\alpha/d}\big(\Lambda_{s_j}[h_j \vee \eta_j]\big)\\ &\le \prod_{j=1}^{d} \sum_{s_j=1}^{\infty} \nu_d^{\alpha/d}\big(\Lambda_{s_j}[h_j \vee \eta_j]\big) \le L.\end{aligned}$$

Thus we obtain that $\vec{h}, \vec{\eta} \in \mathbb{H}_d(d^{-1}\alpha, 2^{-1}L^{1/d})$ implies $\vec{h} \vee \vec{\eta} \in \mathbb{H}_d(\alpha, L)$. Putting $\varkappa = \alpha/d$ and $\mathfrak{L} = 2^{-1}L^{1/d}$ we come to the assertion of the lemma since $\vec{h}, \vec{\eta} \in \mathcal{B}(\mathcal{A})$ implies $\vec{h} \vee \vec{\eta} \in \mathcal{B}(\mathcal{A})$.

A.2. Proof of Lemma 2. Let $\vec{h}, \vec{\eta} \in \mathfrak{S}_d^{\mathrm{const}}$ be fixed. Denote by $\mathcal{J} = \{j = 1, \dots, d : h_j \vee \eta_j = h_j\}$, and suppose first that $\mathcal{J} \ne \varnothing$. Let $\mathcal{J} = \{j_1 < j_2 < \cdots < j_k\}$, $k = |\mathcal{J}|$, and put $\mathcal{J}_l = \{j_1 < j_2 < \cdots < j_l\}$, $l = 1, \dots, k$. Note that for any $x \in \mathbb{R}^d$,

$$\begin{aligned}B_{\vec{h},\vec{\eta}}(x, f) = \bigg|&\int_{\mathbb{R}^d} K_{\vec{h}\vee\vec{\eta}}(t - x)\big[f(t) - f\big(t + \mathbf{E}[\mathcal{J}](x - t)\big)\big]\nu_d(\mathrm{d}t)\\ &- \int_{\mathbb{R}^d} K_{\vec{\eta}}(t - x)\big[f(t) - f\big(t + \mathbf{E}[\mathcal{J}](x - t)\big)\big]\nu_d(\mathrm{d}t)\bigg|.\end{aligned}$$

Here we have used Assumption 1(ii) and $\int \mathcal{K} = 1$. Note also that

$$\begin{aligned}&f(t) - f\big(t + \mathbf{E}[\mathcal{J}](x - t)\big)\\ &\quad= \sum_{l=1}^{k} f\big(t + \mathbf{E}[\mathcal{J}_{l-1}](x - t)\big) - f\big(t + \mathbf{E}[\mathcal{J}_l](x - t)\big),\end{aligned} \tag{A.3}$$

where we have put $\mathcal{J}_0 = \varnothing$. Thus we have for any $x \in \mathbb{R}^d$,

$$
\begin{aligned}
B_{\vec{h},\vec{\eta}}(x, f) \le \sum_{l=1}^{k} \biggl| \int_{\mathbb{R}^d} K_{\vec{h}\vee\vec{\eta}}(t-x)\bigl[f\bigl(t + \mathbf{E}[\mathcal{J}_{l-1}](x-t)\bigr) \\
- f\bigl(t + \mathbf{E}[\mathcal{J}_l](x-t)\bigr)\bigr]\nu_d(\mathrm{d}t)\biggr| \\
+ \sum_{l=1}^{k} \biggl| \int_{\mathbb{R}^d} K_{\vec{\eta}}(t-x)\bigl[f\bigl(t + \mathbf{E}[\mathcal{J}_{l-1}](x-t)\bigr) \\
- f\bigl(t + \mathbf{E}[\mathcal{J}_l](x-t)\bigr)\bigr]\nu_d(\mathrm{d}t)\biggr|.
\end{aligned} \tag{A.4}
$$

Noting that $h_{j_l} \vee \eta_{j_l} = h_{j_l}$ for any $l = 1, \dots, k$, in view of the definition of $\mathcal{J}$, we have

$$
\begin{aligned}
&\biggl| \int_{\mathbb{R}^d} K_{\vec{h}\vee\vec{\eta}}(t-x)\bigl[f\bigl(t + \mathbf{E}[\mathcal{J}_{l-1}](x-t)\bigr) - f\bigl(t + \mathbf{E}[\mathcal{J}_l](x-t)\bigr)\bigr]\nu_d(\mathrm{d}t)\biggr| \\
&\quad \le \int_{\mathbb{R}^{|\bar{\mathcal{J}}_l|}} |K_{\vec{h}\vee\vec{\eta},\bar{\mathcal{J}}_l}(t_{\bar{\mathcal{J}}_l} - x_{\bar{\mathcal{J}}_l})| b_{\vec{h},j_l}(x_{\mathcal{J}_l}, t_{\bar{\mathcal{J}}_l}) \nu_{|\bar{\mathcal{J}}_l|}(\mathrm{d}t_{\bar{\mathcal{J}}_l}) \\
&\quad = \bigl[|K_{\vec{h}\vee\vec{\eta}}| \star b_{\vec{h},j_l}\bigr]_{\mathcal{J}_l}(x), \\
&\biggl| \int_{\mathbb{R}^d} K_{\vec{\eta}}(t-x)\bigl[f\bigl(t + \mathbf{E}[\mathcal{J}_{l-1}](x-t)\bigr) - f\bigl(t + \mathbf{E}[\mathcal{J}_l](x-t)\bigr)\bigr]\nu_d(\mathrm{d}t)\biggr| \\
&\quad \le \int_{\mathbb{R}^{|\bar{\mathcal{J}}_l|}} |K_{\vec{\eta},\bar{\mathcal{J}}_l}(t_{\bar{\mathcal{J}}_l} - x_{\bar{\mathcal{J}}_l})| b_{\vec{h},j_l}(x_{\mathcal{J}_l}, t_{\bar{\mathcal{J}}_l}) \nu_{|\bar{\mathcal{J}}_l|}(\mathrm{d}t_{\bar{\mathcal{J}}_l}) \\
&\quad = \bigl[|K_{\vec{\eta}}| \star b_{\vec{h},j_l}\bigr]_{\mathcal{J}_l}(x).
\end{aligned} \tag{A.5}
$$

Here we have used once again Assumption 1(ii) and $\int \mathcal{K} = 1$.

Thus, for any $\vec{h}, \vec{\eta} \in \mathfrak{S}_d^{\mathrm{const}}$ for which $\mathcal{J} \neq \varnothing$, the first assertion of the lemma follows from (A.4), (A.5) and (A.6). It remains to note that $B_{\vec{h},\vec{\eta}}(\cdot, f) \equiv 0$ if $\mathcal{J} = \varnothing$, and therefore, the first assertion is true with an arbitrary choice of $\{j_1, \dots, j_k\}$. In particular, one can choose $k = d$, which corresponds to $\{j_1, \dots, j_k\} = \{1, \dots, d\}$.

To get the second assertion we choose $\mathcal{J} := \{j_1, \dots, j_k\} = \{1, \dots, d\}$, which yields $\mathcal{J}_l = \{1, \dots, l\}$, and we note that (A.3) remains true. Repeating these computations leads to (A.6), with $\vec{\eta}$ replaced by $\vec{h}$, and we come to the second assertion of the lemma.

A.3. Proof of Lemma 5. As previously mentioned, if $r^*(s) = s$, the assertion of the lemma is proved in Nikol'skiĭ (1977), Section 6.9. Thus it remains to study the case $r^* > s$, where we put $r^* = \max_{j=1,\dots,d} r_j$. Set also $\vec{r}^* = (r^*, \dots, r^*)$, and denote $J_+ = \{j : r_j \ge s\}$ and $J_- = \{1, \dots, d\} \setminus J_+$.

The assumption $\tau(r^*(s)) = \tau(r^*) > 0$, together with $r_j \le r^*$ for any $j = 1, \ldots, d$, makes possible the application of the theorem of Section 6.9 in Nikol'skiĭ (1977), which yields

$$\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \subseteq \mathbb{N}_{\vec{r}^*,d}(\vec{\boldsymbol{\gamma}}(r^*), \mathbf{c}\vec{L}). \tag{A.6}$$

Note that for any $j \in J_-$ we have $\|f\|_{r_j} \le L_j$ since $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ and $\|f\|_{r^*} \le \mathbf{c}L_j$ in view of (A.6). Noting that $r_j < s = r_j(s) < r^*$ we have $\|f\|_{r_j(s)} \le \mathbf{c}_1 L_j$ for any $j \in \mathcal{J}_-$ in view of Hölder's inequality. It remains to note that $r_j(s) = r_j$ for any $j \in J_+$, and we assert that

$$\|f\|_{r_j(s)} \le \mathbf{c}_1 L_j \qquad \forall j = 1, \ldots, d. \tag{A.7}$$

Since $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ and $\boldsymbol{\gamma}_j(s) = \beta_j$, $r_j(s) = r_j$, $j \in \mathcal{J}_+$, one has

$$\begin{aligned} \|\Delta_{u,j}^{k_j} g\|_{r_j(s),\mathbb{R}^d} = \|\Delta_{u,j}^{k_j} g\|_{r_j,\mathbb{R}^d} \le L_j |u|^{\beta_j} = L_j |u|^{\boldsymbol{\gamma}_j(s)} \\ \forall u \in \mathbb{R}, \forall j \in \mathcal{J}_+. \end{aligned} \tag{A.8}$$

Let now $j \in J_-$. If $r^* = \infty$, we have

$$\begin{aligned} \|\Delta_{u,j}^{k_j} g\|_{s,\mathbb{R}^d}^s &\le \|\Delta_{u,j}^{k_j} g\|_{r_j,\mathbb{R}^d}^{r_j} \|\Delta_{u,j}^{k_j} g\|_{\infty,\mathbb{R}^d}^{s-r_j} \\ &\le \mathbf{c}^{s-r_j} L_j^s |u|^{r_j\beta_j + (s-r_j)\beta_j \tau(\infty)\tau^{-1}(r_j)}, \end{aligned} \tag{A.9}$$

in view of (A.6). If $r^* < \infty$, writing

$$s = \frac{r_j(r^* - s)}{r^* - r_j} + \frac{r^*(s - r_j)}{r^* - r_j}$$

and applying the Hölder inequality with exponents $\frac{r^*-r_j}{r^*-s}$ and $\frac{r^*-r_j}{s-r_j}$, we obtain

$$\begin{aligned} \|\Delta_{u,j}^{k_j} g\|_{s,\mathbb{R}^d}^s &\le \left(\|\Delta_{u,j}^{k_j} g\|_{r_j,\mathbb{R}^d}\right)^{(r^*-s)r_j/(r^*-r_j)} \left(\|\Delta_{u,j}^{k_j} g\|_{r^*,\mathbb{R}^d}\right)^{(s-r_j)r^*/(r^*-r_j)} \\ &\le \mathbf{c}_1^{(s-r_j)r^*/(r^*-r_j)} L_j^s |u|^{a_j} \qquad \forall u \in \mathbb{R}, \end{aligned} \tag{A.10}$$

in view of (A.6) with

$$a_j = \frac{(r^* - s)\beta_j r_j}{r^* - r_j} + \frac{\boldsymbol{\gamma}_j(r^*)(s - r_j)r^*}{r^* - r_j} = \frac{(r^* - s)\beta_j r_j}{(r^* - r_j)} + \frac{\tau(r^*)(s - r_j)\beta_j r^*}{\tau(r_j)(r^* - r_j)}.$$

Note that (A.9) is a particular case of (A.10).

We easily compute that $b_j := \tau(r_j)(r^* - s)\beta_j r_j + \tau(r^*)(s - r_j)\beta_j r^* = s\beta_j \tau(s)(r^* - r_j)$ and therefore,

$$a_j := \frac{b_j}{\tau(r_j)(r^* - r_j)} = \frac{s\tau(s)\beta_j}{\tau(r_j)} = s\boldsymbol{\gamma}_j(s).$$

Thus we obtain from (A.6) that

$$\|\Delta_{u,j}^{k_j} g\|_{s,\mathbb{R}^d} \le \mathbf{c}_1 L_j^s |u|^{\boldsymbol{\gamma}_j(s)} \qquad \forall u \in \mathbb{R}, \forall j \in \mathcal{J}_-. \tag{A.11}$$

The required embedding follows now from (A.7), (A.8) and (A.11).

A.4. Proof of Lemma 6. We obviously have

$$b_{\vec{\mathrm{h}},j}(x) = \sup_{\eta\le \mathrm{h}_j,\eta\in\mathfrak{H}}\left|\int_{\mathbb{R}} w_\ell(\mathfrak{z})\big[f(x+\mathfrak{z}\eta\mathbf{e}_j)-f(x)\big]\nu_1(\mathrm{d}\mathfrak{z})\right|$$
$$= \sup_{\eta\le \mathrm{h}_j,\eta\in\mathfrak{H}}\left|\int_{\mathbb{R}} w_\ell(\mathfrak{z})\big[\Delta_{\mathfrak{z}\eta,j}f(x)\big]\nu_1(\mathrm{d}\mathfrak{z})\right|.$$

For $j=1,\dots,d$ we have

$$\int_{\mathbb{R}} w_\ell(\mathfrak{z})\Delta_{\mathfrak{z}\eta,j}f(x)\nu_1(\mathrm{d}\mathfrak{z})$$
$$= \int_{\mathbb{R}}\sum_{i=1}^{\ell}\binom{\ell}{i}(-1)^{i+1}\frac{1}{i}w\left(\frac{\mathfrak{z}}{i}\right)\big[\Delta_{\eta\mathfrak{z},j}f(x)\big]\nu_1(\mathrm{d}\mathfrak{z})$$
$$= (-1)^{\ell-1}\int_{\mathbb{R}} w(z)\sum_{i=1}^{\ell}\binom{\ell}{i}(-1)^{i+\ell}\big[\Delta_{iz\eta,j}f(x)\big]\nu_1(\mathrm{d}z)$$
$$= (-1)^{\ell-1}\int_{\mathbb{R}} w(z)\big[\Delta^{\ell}_{z\eta,j}f(x)\big]\nu_1(\mathrm{d}z).$$

The last equality follows from the definition of ℓth order difference operator (3.1). Thus, for any $j=1,\dots,d$ and any $x\in(-b,b)^d$,

$$\begin{aligned} b_{\vec{\mathrm{h}},j}(x,f) &= \sup_{\eta\le \mathrm{h}_j,\eta\in\mathfrak{H}}\left|\int_{\mathbb{R}} w(z)\big[\Delta^{\ell}_{z\eta,j}f(x)\big]\nu_1(\mathrm{d}z)\right| \\ &\le \sum_{\eta\le \mathrm{h}_j}\left|\int_{\mathbb{R}} w(z)\big[\Delta^{\ell}_{z\eta,j}f(x)\big]\nu_1(\mathrm{d}z)\right|, \end{aligned} \tag{A.12}$$

since $\mathfrak{H}$ is a discrete set. Therefore, by the Minkowski inequality for integrals [see, e.g., Folland (1999), Section 6.3] and the triangle inequality, choosing $\mathfrak{s}$ from the relation $e^{-\mathfrak{s}-2}=\mathrm{h}_j$ (recall that $\mathrm{h}_j\in\mathfrak{H}$), we obtain

$$\|b_{\vec{\mathrm{h}},j}(\cdot,f)\|_{\mathbf{r},\mathbb{R}^d} \le \sum_{s=\mathfrak{s}}^{\infty}\int_{-1/(2\ell)}^{1/(2\ell)}|w(z)|\|\Delta^{\ell}_{ze^{-s-2},j}f\|_{\mathbf{r},\mathbb{R}^d}\nu_1(\mathrm{d}z).$$

Here we have also used that w is compactly supported on $[-1/(2\ell),1/(2\ell)]$.

Note that $\Delta^{\ell}_{ze^{-s-2},j}f$ is supported on $\mathcal{Y}:=(-b-1/2,b+1/2)^d$ for any $z\in[-1/(2\ell),1/(2\ell)]$. Hence, taking into account that $\mathbf{r}\le r_j$, we get

$$\begin{aligned}\|\Delta^{\ell}_{ze^{-s-2},j}f\|_{\mathbf{r},\mathbb{R}^d} &= \|\Delta^{\ell}_{ze^{-s-2},j}f\|_{\mathbf{r},\mathcal{Y}} \le (2b+1)^{d(1/\mathbf{r}-1/r_j)}\|\Delta^{\ell}_{ze^{-s-2},j}f\|_{r_j,\mathcal{Y}} \\ &\le (2b+1)^d\|\Delta^{\ell}_{ze^{-s-2},j}f\|_{r_j,\mathbb{R}^d} \le (2b+1)M_j\big(ze^{-s-2}\big)^{\beta_j},\end{aligned}$$

since $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{M})$. Hence, for any $\mathbf{r} \in [1, r_j]$,

$$\|b_{\vec{\mathrm{h}},j}(\cdot, f)\|_{\mathbf{r},\mathbb{R}^d} \le (2b+1)^d M_j \int_{-1/(2\ell)}^{1/(2\ell)} |w(z)||z|^{\beta_j} \nu_1(\mathrm{d}z) \sum_{s=\mathfrak{s}}^{\infty} (e^{-s-2})^{\beta_j}$$
$$\le (2b+1)^d \|w\|_{1,\mathbb{R}^d} (1 - e^{-\beta_j})^{-1} M_j \mathrm{h}_j^{\beta_j}.$$

This proves (7.5).

The inequality in (7.6) follows by the same reasoning, with r_j replaced by q_j, β_j replaced by γ_j and with the use of embedding (7.4).

A.5. Proof of Lemma 7. We will analyze the set $\mathfrak{H}_\varepsilon(\vartheta, p)$ separately for different values of (ϑ, p).

1^0. Case $\varkappa(p) > 0$. If $\varkappa(p^*) \ge 0$, we have $r_j \le p^* \le \omega(2 + 1/\beta)$ for all $j = 1, \dots, d$. Therefore, for any $m \ge 0$,

$$\widetilde{\boldsymbol{\eta}}_j(m) \le e^{-2}(L_j^{-1}\varphi)^{1/\beta_j} \qquad \forall j = 1, \dots, d.$$

Thus, for all $\varepsilon > 0$ small enough, $\bar{\boldsymbol{\eta}}_j(m) := \widetilde{\boldsymbol{\eta}}_j(m) < \mathfrak{h}_\varepsilon$. This yields

$$\text{(A.13)} \qquad \mathfrak{h}_{s_j(m)} \le \bar{\boldsymbol{\eta}}_j(m) < e\mathfrak{h}_{s_j(m)}, \qquad j = 1, \dots, d.$$

If $\varkappa(p^*) < 0$, which is possible only if $p^* > p$ in view of $\varkappa(p) > 0$, we have for any $0 \le m \le \widetilde{\mathbf{m}}$ and any $j = 1, \dots, d$,

$$\widetilde{\boldsymbol{\eta}}_j(m) \le e^{-2}(L_j^{-1}\varphi)^{1/\beta_j} e^{2dm(1/\beta_j - \omega(2+1/\beta)/(\beta_j p^*))}$$
$$\le e^{-2}(L_j^{-1}\varphi e^{2d\widetilde{\mathbf{m}}(-\varkappa(p^*)/p^*)})^{1/\beta_j}.$$

The definition of $\widetilde{\mathbf{m}}$ implies $L_j^{-1}\varphi e^{2d\widetilde{\mathbf{m}}(-\varkappa(p^*)/p^*)} \le \mathfrak{h}_\varepsilon^\ell$, and we assert that for all $\varepsilon > 0$ small enough,

$$\bar{\boldsymbol{\eta}}_j(m) := \widetilde{\boldsymbol{\eta}}_j(m) < \mathfrak{h}_\varepsilon^{\ell/\beta_j} \le \mathfrak{h}_\varepsilon$$

since $\beta_j \le \ell$ and $\mathfrak{h}_\varepsilon < 1$. Thus we conclude that for any $(\vartheta, p) \in \mathcal{P}^{\mathrm{consist}}$ provided $\varkappa(p) > 0$ and for all $\varepsilon > 0$ small enough,

$$\text{(A.14)} \qquad \mathfrak{H}_\varepsilon(\vartheta, p) \subset \mathfrak{H}^d(\mathfrak{h}_\varepsilon).$$

This implies, in particular, that (A.13) takes place when $\varkappa(p^*) < 0$ as well. Hence we obtain

$$\text{(A.15)} \quad e^{-d} \prod_{j=1}^d \mathfrak{h}_{s_j(m)} \le \prod_{j=1}^d \bar{\boldsymbol{\eta}}_j(m) = e^{-2d} L_\beta^{-1} \varphi^{1/\beta} e^{-4dm} \le \prod_{j=1}^d \mathfrak{h}_{s_j(m)}.$$

This yields

$$\text{(A.16)} \qquad \mathbf{s}(m) \ne \mathbf{s}(n) \qquad \forall m \ne n, m, n = 0, \dots, \widetilde{\mathbf{m}}.$$

2^0. Case $\varkappa(p) \le 0$, $\tau(p^*) \le 0$. Since we consider $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$, the later case is possible only if $p^* > p$. This implies $\varkappa(p^*) < 0$, and as previously we have

$$\widetilde{\boldsymbol{\eta}}_j(m) \le e^{-2}\big(L_j^{-1}\varphi e^{2d\widetilde{\mathbf{m}}(1-\omega(2+1/\beta)/p^*)}\big)^{1/\beta_j}.$$

This yields, in view of the definition of $\widetilde{\mathbf{m}}$,

$$\widetilde{\boldsymbol{\eta}}_j(m) \le e^{-2}\big((L_0^{-1}\varphi)e^{-2d\widetilde{\mathbf{m}}(\varkappa(p^*)/p^*)}\big)^{1/\beta_j} \le e^{-2}.$$

We conclude that for any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ such that $\varkappa(p) \le 0$ and $\tau(p^*) \le 0$ for all $\varepsilon > 0$,

$$\mathfrak{H}_\varepsilon(\vartheta, p) \subset \mathfrak{H}^d, \tag{A.17}$$

and that (A.13) holds. Hence, in view of (A.15) we assert that (A.16) is also fulfilled.

3^0. Case $\varkappa(p) \le 0$, $\tau(p^*) > 0$. Recall that this case is possible only if $p > 2$, which implies in particular that $\tau(2) > 0$.

We start with presenting some relations between the parameters β, γ, ω and υ whose proofs are given in Section A.8.

$$\gamma < \beta, \qquad \upsilon(2 + 1/\gamma) > p_\pm; \tag{A.18}$$

$$1/\omega - 1/\upsilon = \beta(1/\gamma - 1/\beta)(1 - 1/\omega). \tag{A.19}$$

We deduce from equality (A.19) that

$$\begin{aligned} &\upsilon(2 + 1/\gamma) - \omega(2 + 1/\beta) \\ &\quad = \omega\upsilon\big[(2 + 1/\beta)(1/\omega - 1/\upsilon) + (1/\gamma - 1/\beta)\omega^{-1}\big] \\ &\quad = 2\beta\tau(2)\omega\upsilon(1/\gamma - 1/\beta). \end{aligned} \tag{A.20}$$

Using (A.20) we easily get that for any $\mathbf{r} > 0$,

$$1 - \frac{\mathbf{r} - \upsilon(2 + 1/\gamma)}{2\mathbf{r}\beta\omega\tau(2)} - \frac{(1/\gamma - 1/\beta)\upsilon}{\mathbf{r}} = \frac{(2 + 1/\beta)\tau(\mathbf{r})}{2\tau(2)}. \tag{A.21}$$

Since $\upsilon(2 + 1/\gamma) \ge p_\pm$, $\widetilde{\mathbf{m}} > \widehat{\mathbf{m}}$, in view of the definition of $\widetilde{\mathbf{m}}$ and $q_j \le p_\pm$, $j \in \mathcal{J}_\pm$ and $q_j = \infty$, $j \in \mathcal{J}_\infty$, we get

$$\begin{aligned} &\bar{\boldsymbol{\eta}}_j(m) = \widetilde{\boldsymbol{\eta}}_j(m) \le e^{-2}\big(L_j^{-1}\varphi\big)^{1/\beta_j} e^{2d\widehat{\mathbf{m}}(1/\beta_j - \omega(2+1/\beta)/(\beta_j p^*))}, \qquad m \le \widehat{\mathbf{m}}; \\ &\bar{\boldsymbol{\eta}}_j(m) = \widehat{\boldsymbol{\eta}}_j(m) \le e^{-2}\big(L_j^{-1}\varphi\big)^{1/\gamma_j} e^{2d\widehat{\mathbf{m}}(1/\gamma_j - \upsilon(2+1/\gamma)/(\gamma_j q_j))}\bigg[\frac{L_\gamma\varphi^{1/\beta}}{L_\beta\varphi^{1/\gamma}}\bigg]^{\upsilon/(\gamma_j q_j)}, \\ &\qquad\qquad m > \widehat{\mathbf{m}}. \end{aligned} \tag{A.22}$$

We obtain, in view of the definition of $\widehat{\mathbf{m}}$,

$$\{\bar{\boldsymbol{\eta}}_j(m)\}^{\beta_j} \le e^{-2\beta_j} L_j^{-1} \varphi^{(2+1/\beta)\tau(p^*)/(2\tau(2))}, \qquad m \le \widehat{\mathbf{m}};$$

$$\{\bar{\boldsymbol{\eta}}_j(m)\}^{\gamma_j} \le e^{-2\gamma_j} T_1 L_j^{-1} \varphi^{1-(q_j-\upsilon(2+1/\gamma))/(2q_j\beta\omega\tau(2))-(1/\gamma-1/\beta)\upsilon/q_j}, \qquad m > \widehat{\mathbf{m}}, j \in \mathcal{J}_{\pm};$$

$$\{\bar{\boldsymbol{\eta}}_j(m)\}^{\gamma_j} \le e^{-2\gamma_j} L_j^{-1} T_2 \varphi^{(2+1/\beta)(1-1/\omega)/(2\tau(2))}, \qquad m > \widehat{\mathbf{m}}, j \in \mathcal{J}_{\infty},$$

where $T_1 = T_1(L_\gamma/L_\beta)$ and $T_2 = T_2(L_\gamma/L_\beta)$ can be easily deduced.

Thus we assert that for any $j = 1, \dots, d$ and any $\varepsilon > 0$ small enough,

$$\bar{\boldsymbol{\eta}}_j(m) \le \mathfrak{h}_\varepsilon \qquad \forall m \le \widehat{\mathbf{m}}. \tag{A.23}$$

Moreover, if $j \in \mathcal{J}_{\pm}$, we have in view of (A.21),

$$\{\bar{\boldsymbol{\eta}}_j(m)\}^{\gamma_j} \le e^{-2\gamma_j} T_1 L_j^{-1} \varphi^{(2+1/\beta)\tau(q_j)/(2\tau(2))} \to 0, \qquad \varepsilon \to 0,$$

since $\tau(q_j) > 0$ for any $j = 1, \dots, d$ in view of $\tau(p^*) > 0$.

Note also that if $\mathcal{J}_\infty \ne \varnothing$, then $p^* = \infty$, and therefore, $\tau(\infty) = 1 - 1/\omega > 0$ and

$$e^{-2\gamma_j} T_2 L_j^{-1} \varphi^{(2+1/\beta)(1-1/\omega)/(2\tau(2))} \to 0, \qquad \varepsilon \to 0.$$

Hence, for all $\varepsilon > 0$ small enough $\bar{\boldsymbol{\eta}}_j(m) \le \mathfrak{h}_\varepsilon, \forall m > \widehat{\mathbf{m}}$. Taking into account (A.23), we conclude that (A.13) and (A.14) hold in the case $\varkappa(p) \le 0, \tau(p^*) > 0$. Moreover, (A.16) is fulfilled if $m \le \widehat{\mathbf{m}}$ as well in view of (A.15).

On the other hand, in view of (A.13),

$$e^{-d} \prod_{j=1}^{d} \mathfrak{h}_{s_j(m)} \le \prod_{j=1}^{d} \bar{\boldsymbol{\eta}}_j(m) = e^{-2d} L_\beta^{-1} \varphi^{1/\beta} e^{-4dm} \le \prod_{j=1}^{d} \mathfrak{h}_{s_j(m)} \qquad \forall m > \widehat{\mathbf{m}}, \tag{A.24}$$

and therefore, (A.16) is fulfilled for any $m \ge 0$.

A.6. Proof of Lemma 8. Let $\mu \in (0, 1)$ be the number we will choose later, and put $\vec{\mu} = (\mu, \dots, \mu)$. Without loss of generality one can assume that $\vec{\mu} \in \mathfrak{S}_d^{\mathrm{const}}$. For any $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ introduce

$$S_{\vec{\mu}}(x, f) = \int_{\mathbb{R}^d} K_{\vec{\mu}}(t - x) f(t) \nu_d(\mathrm{d}t), \qquad x \in \mathbb{R}^d,$$

where we recall that K is given in (3.5).

1^0. Let us prove that for any $\mu \in (0, 1)$,

$$S_{\vec{\mu}}(\cdot, f) \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L}) \qquad \forall f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}). \tag{A.25}$$

First, we note that $S_{\vec{\mu}}(\cdot, f)$ is compactly supported on $(-b-1, b+1)^d$ in view of the definition of the kernel K, since $\mu \in (0, 1)$. Next, taking into account that K is Lipschitz-continuous and compactly supported as well as $f \in \mathbb{L}_{r^*}(\mathbb{R}^d)$, $r^* = \max_{l=1,\dots,d} r_l$, since $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$, and applying Hölder's inequality, we can assert that $S_{\vec{\mu}}(\cdot, f) \in \mathbb{C}(\mathbb{R}^d)$ and moreover

$$S_{\vec{\mu}}(\cdot, f) \in \mathbb{L}_q(\mathbb{R}^d) \qquad \forall q \geq 1. \tag{A.26}$$

Thus $S_{\vec{\mu}}(\cdot, f) \in \mathbb{C}_{\mathbb{K}}(\mathbb{R}^d)$. It remains to prove that $S_{\vec{\mu}}(\cdot, f) \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L})$. Indeed, applying the Young inequality we obtain for any $j = 1, \dots, d$,

$$\|S_{\vec{\mu}}(\cdot, f)\|_{r_j,\mathbb{R}^d} \leq \|K\|_{1,\mathbb{R}^d} \|f\|_{r_j,\mathbb{R}^d} \leq L_j \|K\|_{1,\mathbb{R}^d},$$

since $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$. Moreover, for any $j = 1, \dots, d$, $k \in \mathbb{N}^*$ and any $u \in \mathbb{R}$,

$$\begin{aligned}\Delta_{u,j}^k S_{\vec{\mu}}(x, f) &:= \Delta_{u,j}^k \left\{ \int_{\mathbb{R}^d} K_{\vec{\mu}}(z) f(x+z) \nu_d(\mathrm{d}z) \right\} \\ &= \int_{\mathbb{R}^d} K_{\vec{\mu}}(z) \{\Delta_{u,j}^k f(x+z)\} \nu_d(\mathrm{d}z) \\ &= \int_{\mathbb{R}^d} K_{\vec{\mu}}(t-x) \{\Delta_{u,j}^k f(t)\} \nu_d(\mathrm{d}t).\end{aligned}$$

Thus, applying the Young inequality, we have for any integer $k_j > \beta_j$,

$$\|\Delta_{u,j}^{k_j} S_{\vec{\mu}}(\cdot, f)\|_{r_j,\mathbb{R}^d} \leq \|K\|_{1,\mathbb{R}^d} \|\Delta_{u,j}^{k_j} f\|_{r_j,\mathbb{R}^d} \leq L_j \|K\|_{1,\mathbb{R}^d} |u|^{\beta_j} \qquad \forall u \in \mathbb{R},$$

since $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$. Moreover, $S_{\vec{\mu}}(\cdot, f) \in \mathbb{L}_2(\mathbb{R}^d)$ in view of (A.26).

We conclude that $S_{\vec{\mu}}(\cdot, f) \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \mathbf{a}\vec{L})$, and therefore, (A.25) is established.

2^0. We will need the following auxiliary result. For any $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ there exists $\mathrm{p} > p$ such that

$$f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \quad \Rightarrow \quad f \in \mathbb{L}_{\mathrm{p}}(\mathbb{R}^d). \tag{A.27}$$

Indeed, if $p^* > p$, we can choose $\mathrm{p} = p^*$ in view of the definition of an anisotropic Nikol'skii class. If $p < 2$, one can choose $\mathrm{p} = 2$ since the definition of $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ implies that $f \in \mathbb{L}_2(\mathbb{R}^d)$.

It remains to consider the cases $p^* = p$ and $p \geq 2$. Since $(\vartheta, p) \in \mathcal{P}^{\text{consist}}$ necessarily in this case $\tau(p) > 0$, and therefore, one can find $\mathrm{p} > p$ such that $\tau(\mathrm{p}) > 0$. In view of $p^* = p < \mathrm{p}$ and $\tau(\mathrm{p}) > 0$, the assertion of Lemma 5 holds with $s = \mathrm{p}$ and $\vec{r}(s) = (\mathrm{p}, \dots, \mathrm{p})$, and therefore, $f \in \mathbb{L}_{\mathrm{p}}(\mathbb{R}^d)$ in view of the definition of an anisotropic Nikol'skii class. Thus (A.27) is established.

3^0. Let $f, g \in \mathbb{L}_p(\mathbb{R}^d)$ be arbitrary functions. We obviously have

$$\sup_{\vec{h} \in \mathfrak{S}_d} |\mathcal{B}_{\vec{h}}^{(p)}(g) - \mathcal{B}_{\vec{h}}^{(p)}(f)| \leq 3 \sup_{\vec{\mathrm{h}} \in \mathfrak{S}_d} \|S_{\vec{\mathrm{h}}}(\cdot, g - f)\|_p + \|g - f\|_p.$$

Since K is compactly supported on $[-1/2, 1/2]^d$, we obviously have that

$$|S_{\vec{\mathrm{h}}}(x, g - f)| \leq \|K\|_{\infty,\mathbb{R}^d} M[|g - f|](x), \qquad x \in \mathbb{R}^d.$$

Applying (p, p)-strong maximal inequality (7.8), we obtain for any $p > 1$,

$$\|S_{\vec{\mathrm{h}}}(\cdot, g - f)\|_p \le \bar{C}(p)\|K\|_{\infty,\mathbb{R}^d}\|g - f\|_{p,\mathbb{R}^d}.$$

Noting that the right-hand side of the latter inequality is independent of $\vec{\mathrm{h}}$, we obtain finally

$$\sup_{\vec{h}\in\mathfrak{S}_d} |\mathcal{B}^{(p)}_{\vec{h}}(g) - \mathcal{B}^{(p)}_{\vec{h}}(f)| \le \big(3\bar{C}(p)\|K\|_{\infty,\mathbb{R}^d} + 1\big)\|g - f\|_{p,\mathbb{R}^d}.$$

Choosing $g = S_{\vec{\mu}}(\cdot, f)$ and noting that $|S_{\vec{\mu}}(\cdot, f) - f(\cdot)| =: B_{\vec{\mu}}(\cdot, f)$, we get

$$\begin{aligned} &\sup_{\vec{h}\in\mathfrak{S}_d} |\mathcal{B}^{(p)}_{\vec{h}}\big(S_{\vec{\mu}}(\cdot, f)\big) - \mathcal{B}^{(p)}_{\vec{h}}(f)| \\ &\quad\le \big(3\bar{C}(p)\|K\|_{\infty,\mathbb{R}^d} + 1\big)\|B_{\vec{\mu}}(\cdot, f)\|_{p,\mathbb{R}^d}. \end{aligned} \tag{A.28}$$

4^0. Some remarks are in order. First, $B_{\vec{\mu}}(\cdot, f)$ is compactly supported on $\mathbb{K}$ for any $\mu \in (0, 1)$. Next, $B_{\vec{\mu}}(\cdot, f) \in \mathbb{L}_{\mathrm{p}}(\mathbb{R}^d)$ in view of (A.26) and (A.27). Finally, in view of (5.6) and the first assertion of Lemma 6, we have $\limsup_{\mu\to 0} \|B_{\vec{\mu}}(\cdot, f)\|_{1,\mathbb{R}^d} = 0$.

The above allows us to assert that $\limsup_{\mu\to 0} \|B_{\vec{\mu}}(\cdot, f)\|_{p,\mathbb{R}^d} = 0$. This yields, together with (A.28), that for any $\varkappa > 0$ and any $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$, one can find $\boldsymbol{\mu} = \boldsymbol{\mu}(\varkappa, f)$ such that

$$\sup_{\vec{h}\in\mathfrak{S}_d} |\mathcal{B}^{(p)}_{\vec{h}}\big(S_{\vec{\boldsymbol{\mu}}}(\cdot, f)\big) - \mathcal{B}^{(p)}_{\vec{h}}(f)| \le \varkappa,$$

where as previously $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu}, \dots, \boldsymbol{\mu})$.

This obviously implies, for any $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$ and any $\mathbb{H} \subseteq \mathfrak{S}_d$,

$$\begin{aligned} \inf_{\vec{h}\in\mathbb{H}} \big[\mathcal{B}^{(p)}_{\vec{h}}(f) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\big] &\le \inf_{\vec{h}\in\mathbb{H}} \big[\mathcal{B}^{(p)}_{\vec{h}}\big(S_{\vec{\boldsymbol{\mu}}}(\cdot, f)\big) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\big] + \varkappa \\ &\le \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \inf_{\vec{h}\in\mathbb{H}} \big[\mathcal{B}^{(p)}_{\vec{h}}(g) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\big] + \varkappa, \end{aligned}$$

where to get the last inequality we have used (A.25). Since the right-hand side of the latter inequality is independent of f, one gets

$$\sup_{f\in\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})} \inf_{\vec{h}\in\mathbb{H}} \big[\mathcal{B}^{(p)}_{\vec{h}}(f) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\big] \le \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \inf_{\vec{h}\in\mathbb{H}} \big[\mathcal{B}^{(p)}_{\vec{h}}(g) + \varepsilon\Psi_{\varepsilon,p}(\vec{h})\big] + \varkappa,$$

and the assertion of the lemma follows since $\varkappa$ is an arbitrary number.

A.7. Proof of Lemma 9. We obviously have $U_{\vartheta,p}(x,g) \le b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(0)}}(x,g) + \varpi_\varepsilon V_{\mathbf{s}(0)}^{-1/2}$ and therefore,

$$U := \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \|U_{\vartheta,p}(\cdot,g)\|_{p^*} \le \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \|b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(0)}}(x,g)\|_{p^*} + (2b)^{d/p}\varpi_\varepsilon V_{\mathbf{s}(0)}^{-1/2}.$$

Note that in view of (7.28) $\varpi_\varepsilon V_{\mathbf{s}(0)}^{-1/2} \to 0$, $\varepsilon \to 0$ and therefore, for all $\varepsilon > 0$ small enough,

$$U \le \Upsilon_1 \sup_{g\in\mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})} \|b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(0)}}(x,g)\|_{p^*} + L^*. \tag{A.29}$$

Recall that $b^*_{\vec{h}}(x,g) = \sup_{J\in\mathfrak{J}} \sup_{j=1,\dots,d} M_J[b_{\vec{h},j}](x)$, and therefore, we obtain first, applying (7.9),

$$\|b^*_{\vec{\mathfrak{h}}_{\mathbf{s}(0)}}(\cdot,g)\|_{p^*} \le 2^d \mathbf{C}_{p^*} \sum_{j=1}^{d} \|b_{\vec{\mathfrak{h}}_{\mathbf{s}(0)},j}\|_{p^*}. \tag{A.30}$$

Next, we have for any $j = 1,\dots,d$ and any $x \in \mathbb{R}^d$,

$$\begin{aligned} b_{\vec{\mathfrak{h}}_{\mathbf{s}(0)},j}(x) &:= \sup_{k:\,\mathfrak{h}_k \le \mathfrak{h}_{s_j(0)}} \left| \int_{\mathbb{R}} w_\ell(u) g(x + u\mathfrak{h}_k \mathbf{e}_j)\nu_1(\mathrm{d}u) - g(x) \right| \\ &\le \|w_\ell\|_{\infty,\mathbb{R}^d} M_{J_j}[g](x) + |g(x)|, \end{aligned}$$

where we have denoted $J_j = \{1,\dots,d\} \setminus \{j\}$. Thus, applying once again (7.9) we obtain

$$\|b_{\vec{\mathfrak{h}}_{\mathbf{s}(0)},j}\|_{p^*} \le \big(\mathbf{C}_{p^*}\|w_\ell\|_{\infty,\mathbb{R}^d} + 1\big)\|g\|_{p^*}.$$

Noting that in view of the definition of the Nikol'skii class, $\|g\|_{p^*} \le \mathbf{a}L^*$ for any $g \in \mathbb{N}^*_{\vec{r},d}(\vec{\beta},\mathbf{a}\vec{L})$, and the assertion of the lemma follows from (A.29) and (A.30).

A.8. Proof of formulas (A.18) and (A.19). In view of the definition of $\vec{\gamma}$ we have

$$\frac{1}{\gamma} = \sum_{j\in J_\pm} \frac{\tau(r_j)}{\tau(p_\pm)\beta_j} + \sum_{j\in J_\infty} \frac{1}{\beta_j} \ge \frac{1}{\beta}$$

since $\tau(r_j) \ge \tau(p_\pm)$. In view of the definition υ,

$$\frac{p_\pm}{\upsilon} = \sum_{j\in J_\pm} \frac{1}{\gamma_j} \le \sum_{j\in J_\pm} \frac{1}{\gamma_j} + \sum_{j\in J_\infty} \frac{1}{\beta_j} = \frac{1}{\gamma}$$

and therefore, $p_\pm \le \upsilon/\gamma < \upsilon(2 + 1/\gamma)$.

Proof of (A.19). First, we remark that

$$p_{\pm}\left(\frac{1}{\omega}-\frac{1}{\upsilon}\right)+\frac{1}{\gamma}-\frac{1}{\beta}=\sum_{j\in J_{\pm}}\left(\left[\frac{p_{\pm}}{r_j\beta_j}-\frac{1}{\gamma_j}\right]+\left[\frac{1}{\gamma_j}-\frac{1}{\beta_j}\right]\right)$$
$$=p_{\pm}\sum_{j\in J_{\pm}}\left(\frac{1}{r_j\beta_j}-\frac{1}{p_{\pm}\beta_j}\right)=:Ap_{\pm}.$$

Next,

$$\begin{aligned}\sum_{j\in J_{\pm}}\frac{1}{\gamma_j}&=\sum_{j\in J_{\pm}}\frac{\tau_j}{\tau(p_{\pm})\beta_j}=\frac{1}{\tau(p_{\pm})}\sum_{j\in J_{\pm}}\frac{1-1/\omega+1/(r_j\beta)}{\beta_j}\\&=\frac{1-1/\omega}{\tau(p_{\pm})}\sum_{j\in J_{\pm}}\frac{1}{\beta_j}+\frac{1}{\tau(p_{\pm})\beta}\sum_{j\in J_{\pm}}\left(\frac{1}{r_j\beta_j}-\frac{1}{p_{\pm}\beta_j}\right)\\&\quad+\frac{1}{\tau(p_{\pm})\beta p_{\pm}}\sum_{j\in J_{\pm}}\frac{1}{\beta_j}\\&=\sum_{j\in J_{\pm}}\frac{1}{\beta_j}+\frac{A}{\tau(p_{\pm})\beta}.\end{aligned}$$

This yields $\frac{1}{\gamma}-\frac{1}{\beta}=\sum_{j\in J_{\pm}}(\frac{1}{\gamma_j}-\frac{1}{\beta_j})=\frac{A}{\tau(p_{\pm})\beta}$ and therefore,

$$p_{\pm}(1/\omega-1/\upsilon)=(1/\gamma-1/\beta)\big(\tau(p_{\pm})\beta p_{\pm}-1\big)=(1/\gamma-1/\beta)\beta p_{\pm}(1-1/\omega).$$

Relation (A.19) is proved.

REFERENCES

AKAKPO, N. (2012). Adaptation to anisotropy and inhomogeneity via dyadic piecewise polynomial selection. *Math. Methods Statist.* **21** 1–28. MR2901269

BARAUD, Y. and BIRGÉ, L. (2014). Estimating composite functions by model selection. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 285–314. MR3161532

BARAUD, Y., BIRGÉ, L. and SART, M. (2014). A new method for estimation and model selection: ρ-estimation. Available at arXiv:1403.6057v1.

BARAUD, Y., GIRAUD, C. and HUET, S. (2014). Estimator selection in the Gaussian setting. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 1092–1119. MR3224300

BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028

BERTIN, K. (2005). Sharp adaptive estimation in sup-norm for d-dimensional Hölder classes. *Math. Methods Statist.* **14** 267–298. MR2195327

BIRGÉ, L. (2008). Model selection for density estimation with $\mathbb{L}_2$-loss. Available at arXiv:0808.1416v2.

BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946

BUNEA, F., TSYBAKOV, A. B. and WEGKAMP, M. H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35** 1674–1697. MR2351101

CAI, T. T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist.* **27** 898–924. MR1724035

CAI, T. T. and LOW, M. G. (2005). On adaptive estimation of linear functionals. *Ann. Statist.* **33** 2311–2343. MR2211088

CAI, T. T. and LOW, M. G. (2006). Optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **34** 2298–2325. MR2291501

CAVALIER, L. and GOLUBEV, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34** 1653–1677. MR2283712

CAVALIER, L. and TSYBAKOV, A. B. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods Statist.* **10** 247–282. MR1867161

CHICHIGNOUD, M. (2012). Minimax and minimax adaptive estimation in multiplicative regression: Locally Bayesian approach. *Probab. Theory Related Fields* **153** 543–586. MR2948686

CHICHIGNOUD, M. and LEDERER, J. (2014). A robust, adaptive M-estimator for pointwise estimation in heteroscedastic regression. *Bernoulli* **20** 1560–1599. MR3217454

COMTE, F. and LACOUR, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 569–609. MR3088382

DALALYAN, A. and TSYBAKOV, A. B. (2008). Aggregation by exponential weighting, sharp PAC-Bayesian bounds and sparsity. *Mach. Learn.* **72** 39–61.

DE GUZMÁN, M. (1975). *Differentiation of Integrals in R^n. Lecture Notes in Mathematics* **481**. Springer, Berlin. MR0457661

DELYON, B. and JUDITSKY, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3** 215–228. MR1400080

DEVROYE, L. and LUGOSI, G. (1996). A universally acceptable smoothing factor for kernel density estimates. *Ann. Statist.* **24** 2499–2512. MR1425963

DEVROYE, L. and LUGOSI, G. (1997). Nonasymptotic universal smoothing factors, kernel complexity and Yatracos classes. *Ann. Statist.* **25** 2626–2637. MR1604428

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974

EFROĬMOVICH, S. Y. and PINSKER, M. S. (1984). A self-training algorithm for nonparametric filtering. *Automat. Remote Control* **45** 58–65.

EFROĬMOVICH, S. Y. (1986). Nonparametric estimation of the density with unknown smoothness. *Theory Probab. Appl.* **30** 557–568.

EFROMOVICH, S. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **36** 1127–1155. MR2418652

EFROMOVICH, S. and LOW, M. G. (1994). Adaptive estimates of linear functionals. *Probab. Theory Related Fields* **98** 261–275. MR1258989

FOLLAND, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. Wiley, New York. MR1681462

GACH, F., NICKL, R. and SPOKOINY, V. (2013). Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 900–914. MR3112439

GINÉ, E. and NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596. MR2475673

GOLDENSHLUGER, A. (2009). A universal procedure for aggregating estimators. *Ann. Statist.* **37** 542–568. MR2488362

GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. MR2543590

GOLDENSHLUGER, A. and LEPSKI, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields* **143** 41–71. MR2449122

GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214

GOLDENSHLUGER, A. V. and LEPSKI, O. V. (2013). General selection rule from a family of linear estimators. *Theory Probab. Appl.* **57** 209–226. MR3201652

GOLDENSHLUGER, A. and LEPSKI, O. (2014). On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields* **159** 479–543. MR3230001

GOLDENSHLUGER, A. and NEMIROVSKI, A. (1997). Spatial adaptive estimation of smooth nonparametric regression functions. *Math. Methods Statist.* **6** 135–170.

GOLUBEV, G. K. (1992). Nonparametric estimation of smooth densities of a distribution in L_2. *Probl. Inform. Transm.* **1** 52–62.

GOLUBEV, G. K. and NUSSBAUM, M. (1992). Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.* **37** 553–560.

HASMINSKII, R. and IBRAGIMOV, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18** 999–1010. MR1062695

HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35** 2589–2619. MR2382659

HRISTACHE, M., JUDITSKY, A., POLZEHL, J. and SPOKOINY, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.* **29** 1537–1566. MR1891738

IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation: Asymptotic Theory. Applications of Mathematics* **16**. Springer, New York. MR0620321

JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.* **33** 1700–1752. MR2166560

JUDITSKY, A. (1997). Wavelet estimators: Adapting to unknown smoothness. *Math. Methods Statist.* **6** 1–25. MR1456644

JUDITSKY, A. and LAMBERT-LACROIX, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10** 187–220. MR2046772

JUDITSKY, A. B., LEPSKI, O. V. and TSYBAKOV, A. B. (2009). Nonparametric estimation of composite functions. *Ann. Statist.* **37** 1360–1404. MR2509077

JUDITSKY, A. and NEMIROVSKI, A. (2000). Functional aggregation for nonparametric regression. *Ann. Statist.* **28** 681–712. MR1792783

KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. MR1863916

KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2008). Nonlinear estimation in anisotropic multiindex denoising. Sparse case. *Theory Probab. Appl.* **52** 58–77.

KERKYACHARIAN, G., NICKL, R. and PICARD, D. (2012). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Related Fields* **153** 363–404. MR2925578

KERKYACHARIAN, G., THANH, M. and PICARD, D. (2011). Localized spherical deconvilution. *Ann. Statist.* **39** 1042–1068.

LEPSKI, O. (2013). Multivariate density estimation under sup-norm loss: Oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. MR3099129

LEPSKI, O. V. (2015). Upper functions for $\mathbb{L}_p$-norm of Gaussian random fields. *Bernoulli*. To appear. Available at arXiv:1311.4996v1.

LEPSKI, O. V. and LEVIT, B. Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.* **7** 123–156. MR1643256

LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. MR1447734

LEPSKI, O. and SERDYUKOVA, N. (2014). Adaptive estimation under single-index constraint in a regression model. *Ann. Statist.* **42** 1–28. MR3161459

LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 459–470.

LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LEPSKIĬ, O. V. (1992a). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Theory Probab. Appl.* **37** 468–481.

LEPSKIĬ, O. V. (1992b). On problems of adaptive estimation in white Gaussian noise. In *Topics in Nonparametric Estimation*. *Adv. Soviet Math.* **12** 87–106. Amer. Math. Soc., Providence, RI. MR1191692

LEUNG, G. and BARRON, A. R. (2006). Information theory and mixing least-squares regressions. *IEEE Trans. Inform. Theory* **52** 3396–3410. MR2242356

MÜLLER, H.-G. and STADTMÜLLER, U. (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15** 182–201. MR0885731

NEMIROVSKI, A. (2000). Topics in nonparametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640

NEMIROVSKIY, A. S. (1985). Nonparametric estimation of smooth regression functions. *Soviet J. Comput. Systems Sci.* **23** 1–11.

NEUMANN, M. H. (2000). Multivariate wavelet thresholding in anisotropic function spaces. *Statist. Sinica* **10** 399–431. MR1769750

NIKOL'SKIĬ, S. M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya*, 2nd ed. Nauka, Moscow. MR0506247

REYNAUD-BOURET, P., RIVOIRARD, V. and TULEAU-MALOT, C. (2011). Adaptive density estimation: A curse of support? *J. Statist. Plann. Inference* **141** 115–139. MR2719482

RIGOLLET, P. (2006). Adaptive density estimation using the blockwise Stein method. *Bernoulli* **12** 351–370. MR2218559

RIGOLLET, P. and TSYBAKOV, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16** 260–280. MR2356821

RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771. MR2816337

SAMAROV, A. and TSYBAKOV, A. (2007). Aggregation of density estimators and dimension reduction. In *Advances in Statistical Modeling and Inference*. *Ser. Biostat.* **3** 233–251. World Scientific, Hackensack, NJ. MR2416118

TSYBAKOV, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469. MR1700239

TSYBAKOV, A. (2003). Optimal rate of aggregation. In *Proc. COLT*. *Lecture Notes in Artificial Intelligence* **2777** 303–313. Springer, New York.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

WEGKAMP, M. (2003). Model selection in nonparametric regression. *Ann. Statist.* **31** 252–273. MR1962506

ZHANG, C.-H. (2005). General empirical Bayes wavelet methods and exactly adaptive minimax estimation. *Ann. Statist.* **33** 54–100. MR2157796

INSTITUT DE MATHÉMATIQUE DE MARSEILLE
AIX-MARSEILLE UNIVERSITÉ
39, RUE F. JOLIOT-CURIE
13453 MARSEILLE
FRANCE
E-MAIL: oleg.lepski@univ-amu.fr